Alma Mater Studiorum · Università di Bologna

School of Science
Department of Physics and Astronomy
Master Degree Programme in Astrophysics and Cosmology

Constraining the impact of redshift interlopers on the two-point and three-point correlation functions

Graduation Thesis

Presented by:
Nicola Principi

Supervisor:
Prof. Michele Moresco

Co-supervisor:
Prof. Federico Marulli
Prof. Sylvain de la Torre

Academic Year 2023/2024
Graduation Date I

A mia madre Teresa,
che ogni giorno mi aiuta a creare
il mio universo
e a Nonna Gianna,
che spero sempre di scorgere
tra le galassie

Abstract

Since the early 2000s, the study of the Large Scale Structure (LSS) of the Universe has become central to cosmology, revealing insights into its expansion and structural evolution. The European Space Agency (ESA) *Euclid* mission promises to significantly advance this field by delivering the most extensive and precise galaxy map to date in optical and infrared bands, extending the current ones by measuring tens of millions of redshifts. To fully leverage *Euclid*'s data for galaxy clustering analyses, two key advancements are required.

Firstly, capturing the complete cosmological signal necessitates moving beyond traditional statistics, like the two-point correlation function (2PCF). In particular, the three-point correlation function (3PCF) is essential for understanding deviations from Gaussian behaviors in the density field. Secondly, achieving extreme precision in cosmological measurements, as anticipated for *Euclid*, demands rigorous control of systematic effects. A major challenge here is the precise estimation of redshifts. At the expected level of signal-to-noise ratio (SNR) in *Euclid* spectra, errors in redshift determinations due to noise or spectral line mismatches (noise and line interlopers) can degrade the 2PCF and 3PCF signals.

In this study, we assess for the first time the impact of interlopers in the measurement of cosmological parameters with the 3PCF. To do that, we derive and test a new expression for the 3PCF estimator that effectively disentangles the cosmological signal from systematic errors. We successfully implement and test new classes and functions in the *CosmoBolognaLib* (CBL) C++ libraries to measure the cross-correlation signal from interlopers for both the 2PCF and 3PCF. Additionally, we assess the impact of redshift interlopers on the 2PCF and 3PCF signals using a galaxy sample from the *Euclid* official simulations (Flagship2), where ad-hoc interlopers have been inserted based on *Euclid* predictions. Our analysis highlights the importance of interlopers' cross-correlations in recovering accurate 2PCF and 3PCF signals using our newly derived estimators. Furthermore, we model the 2PCF and 3PCF signals considering another large set of *Euclid*-like simulations (EuclidLargeMocks from the *PINOCCHIO simulations*), comparing pure samples (without systematics) to those contaminated by interlopers.

Our findings indicate that redshift interlopers significantly bias the 2PCF and 3PCF signals, lowering them by more than 30%, with peaks around 70%. Furthermore, we find

that cross-correlation signals are crucial in recovering the full 2PCF signal, particularly at low redshifts. Regarding cosmological parameter constraints, we observe that the linear growth factor, $f\,\sigma_8$, for catalogues affected by interloper contamination is reduced by more than 25%. Additionally, modeling both 2PCF and 3PCF reveals that the linear bias parameter, b_1, is damped by 20-30% for the contaminated sample.

This work represents the first step in a full self-consistent treatment of realistic data considering in detail systematic effects as expected from the *Euclid* mission, demonstrating that addressing interloper contamination biases is essential. Future studies should focus on developing effective mitigation strategies.

Contents

Introduction

The main goal of cosmology is to understand the nature of the LSS and the history of the evolution of the Universe. This can be achieved, among many other methods, through three main approaches: Cosmic Microwave Background (CMB) studies, weak lensing (particularly using cosmic shear), and clustering. The *Euclid* mission's two main probes for studying the geometry of the universe are weak lensing and clustering. In this work, I will focus on the latter, which involves studying and understanding the matter distribution in the Universe across cosmic time.

To retrieve information on the LSS, useful statistical tools such as the n-point correlation functions are employed. These functions quantify the excess or deficit of probability of finding N galaxies in a given N-point configuration in our Universe compared to a Universe where galaxies are randomly distributed (hence without any clustering). The inflationary theory predicts a primordial almost Gaussian matter density field, and the properties of any Gaussian field can be fully studied with two-point statistics, specifically the 2PCF or its Fourier transform, the power spectrum (PS). However, during the evolution of the universe, non-linear effects due to small-scale gravity alter the initial field, leading to the development of non-Gaussian features. These features cannot be constrained with two-point statistics and require higher-order moments of the matter distribution. The 3PCF, along with its Fourier transform, the bispectrum (BS), is the simplest statistical tool to probe non-Gaussianity.

The first step in performing statistical analysis with the tracers' distribution is to precisely measure their positions. This can introduce various measurement uncertainties, particularly for the redshift coordinate. Besides the effects of the peculiar velocities of the tracers (RSD) and the assumption of a fiducial cosmology for computing comoving distances (AP effect), which have been extensively studied over the last decade, systematic errors can also occur. These errors, known as redshift interlopers, are galaxies whose redshifts are incorrectly estimated due to emission line mismatches. Redshift interlopers will be the main focus of our study. Their presence in surveys affects the measurement of correlation function signals, thus biasing the constraints on cosmological parameters derived from their modeling. Our main objectives are the following.

Firstly, we aim to derive an expression for a 3PCF estimator capable of disentangling the signal of the interlopers and their cross-correlations from that of the pure galaxies.

While this work has already begun for the 2PCF (e.g. Risso et al., in prep.), its extension to three-point statistics is still missing.

We also plan to implement new classes and functions in the CBL (Marulli et al., 2015 [1]), a comprehensive set of *free software* C++/Python libraries designed for cosmological calculations. These new implementations will successfully use the estimators for the cross-correlations of the 2PCF and 3PCF. As we will discuss, these estimators are crucial for studying the redshift systematics effects on clustering.

Furthermore, we aim to constrain the impact of redshift interlopers on both 2PCF and 3PCF signals by comparing measurements from pure catalogues of galaxies with those affected by interloper contamination. We also intend to quantify the importance of estimating the cross-signals of the redshift interlopers in recovering the total signal of the 2PCF and 3PCF.

Finally, we will study the impact of interloper contamination on constraining cosmological parameters through modeling both 2PCF and 3PCF. We will quantify the bias that systematic uncertainties introduce in the estimation of extensively used parameters such as the linear growth factor, $f\,\sigma_8$, and the linear bias, b_1.

The work is organized as follows:

- In Chapter 1 we present an introduction on basic cosmological notions upon which the concordance model of cosmology is built, needed to understand the purposes and methodology of our study.

- In Chapter 2 we give an overview on clustering, one of the two main probes of the *Euclid* mission and the field of cosmology on which we focus in our work. We will concentrate in particular on the definition of the PS, 2PCF, BS and 3PCF.

- In Chapter 3, we describe the methods used for the measurement and the modelling of the statistical tools introduced in the previous Chapter, focusing on the ones used in this study. We also propose a new estimator for the cross-correlations of the 3PCF.

- In Chapter 4 we present and test our newly developed classes and functions to exploit the cross-correlation estimators described in the previous Chapter.

- In Chapter 5 we describe the Flaghip2 (FS2) catalogue and we present our results on the impact of the redshift interlopers' self and cross-correlations on the measurement of the 2PCF and 3PCF.

- In Chapter 6 we describe the EuclidLargeMocks catalogues and we show our work on the effect of the redshift interlopers on cosmological parameters constraints, estimated through the modelling of both 2PCF and 3PCF.

- Eventually, in Chapter 7 we discuss our results and possible future developments of this work.

Chapter 1

Fundamentals of Cosmology

During the last century our conception of the Universe has changed radically: for example only one hundred years ago we strongly believed that the latter was composed exclusively by our galaxy. Nowadays, we think to have a pretty reliable vision on the Universe's evolution and composition, though there is still plenty of shadows and unanswered questions. The goal of this first Chapter is to describe our knowledge on the life stages of the Universe from the *last scattering surface*[1] until today, giving a brief overview of what we know and what we do not understand yet. The model on which the scientific community agrees for describing the evolution of the Universe is the *Strandard Model of Cosmology* also known as ΛCDM, where Λ is the *cosmological constant* and CDM stands for *cold dark matter*, which are believed to be the two most important components of our Universe. Despite it still has some weak points and open questions, for example on the nature of its main components, the ΛCDM model successfully describes the evolution of the cosmos and it is the best guess we have for explaining the *history of everything*.

In this first Chapter, we will begin with giving a brief overview on the instrument used for studying the evolution of the Universe: the theory of General Relativity (GR). After that, we will talk about the main fundamental concepts of cosmology, such as the *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric, the content of the Universe and the Friedmann equations. Then we will present the ΛCDM model and its main components. Eventually, we will talk about the growth of the cosmic structures starting from small perturbations in the homogeneous density field that was the primordial Universe.

[1] It is very difficult to have any direct information from before the last scattering surface, as at that time the Universe was *optically thick* and the photons produced could not travel freely to us without reacting with other particles.

1.1 General Relativity and the field equations

With the theory of GR (1915), German physicist Albert Einstein worked out a general framework of which the physics of that time, the Newtonian physics, is only a particular case, putting down the foundations for the development of the fields of astrophysics and cosmology. Until 1915 Newtonian Physics was the only available gravitational theory, which however failed to explain the cases in which the ratio between the mass and the radius of the studied system in geometrized units ($c = G = 1$) is not negligible. Even though this condition was verified for the majority of the studied systems (terrestrial physics or even whithin the Solar System, with slight exceptions, such as e.g. the precession of the perihelion of Mercury), this is not the case for more extreme scenarios that were not even on the table before the 20th century, where either the radius is very small (e.g. in very compact objects like neutron stars or balck holes) or the mass is very large. The latter situation happens in cosmology, for at cosmological scales the mass contained within large volumes of Universe can be much greater than their radii. This can be easily seen assuming that the mass contained in a volume of Universe of radius R and mean density $\bar{\rho}$ is:

$$M(R) = \frac{4}{3}\pi\bar{\rho}R^3. \tag{1.1}$$

Assuming a $\bar{\rho} = 10^{-26}\text{kg/m}^3$ in Eq.(1.1), we have $M \gtrsim R$ for $R \gtrsim 6\text{Gpc}$, hence at cosmological scales.

The main strength of GR is its background independence: this means that it does not rely on a fixed spacetime background in its formulation. The theory is indeed formulated using a covariant formalism, that makes it independent of the chosen coordinate system and suitable for describing any kind of Universe. In order to treat the topic, some considerations about that formalism have to be made.

From now on we will use the Einstein summation convention. Let us consider a point P in an n-dimensional manifold M, and an n-dimensional vector space Σ tangent to M in P. For describing the position of P we can use the local Cartesian coordinates in Σ, X^i_P with $i = 1, ..., n$. Now, we can switch to arbitrary coordinates x^μ_P using the *Jacobian* e^i_μ, which is defined in this way:

$$e^i_\mu = \left.\frac{\partial X^i_P(x^\mu)}{\partial x^\mu}\right|_{x^\mu = x^\mu_p}. \tag{1.2}$$

This is also called the frame field and relates our set of arbitrary coordinates to the local Cartesian coordinates in any point of the manifold. In this way we can exploit the Cartesian concept of distance $\mathrm{d}s$ between two points x^μ and $x^\mu + \mathrm{d}x^\mu$:

$$\mathrm{d}s^2 = \delta_{ij}\mathrm{d}X^i\mathrm{d}X^j, \tag{1.3}$$

which using Eq.(2.47), becomes:

$$\mathrm{d}s^2 = \delta_{ij}e^i_\mu(x)e^j_\nu(x)\mathrm{d}x^\mu\mathrm{d}x^\nu. \tag{1.4}$$

If we define the metric tensor as:

$$g_{\mu\nu}(x) = \delta_{ij} e^i_\mu(x) e^j_\nu(x), \tag{1.5}$$

then Eq.(1.4) can be rewritten as:

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{1.6}$$

The metric tensor is a symmetric, rank (0,2) tensor field which specifies the local geometric structure of the manifold at each point. Thanks to these instruments, we are now able to define lengths and angles on the manifold (the Universe), and it can be demonstrated that the geodesics equations are:

$$\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta}\dot{x}^\alpha \dot{x}^\beta = 0, \tag{1.7}$$

where $\Gamma^\mu_{\alpha\beta}$ are the *Christoffel symbols*, defined as:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu}(\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\beta\alpha}). \tag{1.8}$$

Furthermore, also the derivatives have to be generalized for an n-dimensional manifold with generic geometry. The covariant derivative of a vector v^ν ($\nu = 1, ..., n$) is defined as follows:

$$\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\alpha_{\mu\nu} v^\alpha. \tag{1.9}$$

In the case of a flat, Euclidean space ($\Gamma^\alpha_{\mu\nu} = 0$), the covariant derivative coincides with the partial one. Therefore, the Christoffel symbols can be seen as the correction to the derivative in the general case of a non-Euclidean metric, for they encode information about how tangent spaces are connected as one moves along a curve on the manifold. Eventually, defining the Riemann tensor as:

$$R^\alpha_{\beta\mu\nu} = \partial_\mu \Gamma^\alpha_{\beta\nu} - \partial_\nu \Gamma^\alpha_{\beta\mu} + \Gamma^\alpha_{\mu\rho}\Gamma^\rho_{\beta\nu} - \Gamma^\alpha_{\nu\rho}\Gamma^\rho_{\beta\mu}, \tag{1.10}$$

it can be demonstrated that:

$$\nabla_\mu \nabla_\nu v^\alpha - \nabla_\nu \nabla_\mu v^\alpha = R^\alpha_{\beta\mu\nu}. \tag{1.11}$$

It is clear that if $R^\alpha_{\beta\mu\nu} = 0$ in a region of the manifold, that region is flat.

GR suggests that the Universe is a four-dimensional manifold that can be generally described using three spatial coordinates (e.g. x, y and z if we want to use the Cartesian ones) and a temporal one (i.e. t). The geometry of this manifold ($g_{\mu\nu}$ and its derivates) depends on the mass and energy contained in it, following the field equations:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.12}$$

where $g_{\mu\nu}$ is the metric tensor, which describes the geometry of the manifold, $R_{\mu\nu}$ and R are respectively the Ricci tensor and the Ricci scalar:

$$R_{\mu\nu} = R^{\rho}_{\mu\nu\rho}, \tag{1.13}$$

$$R = g^{\mu\nu} R_{\mu\nu}, \tag{1.14}$$

which are contractions of the Riemann tensor, therefore derivates of the metric tensor, given the Eqs. (1.10) and (1.8). $T_{\mu\nu}$ is the energy-momentum tensor, which describes the distribution of mass and energy within the manifold and G is the *gravitational constant* ($G \sim 6.67 \times 10^{-11}\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$). Gravitation can be described using Eq.(1.12) in the following way: the distribution of matter and energy ($T_{\mu\nu}$) curves the geometry of the manifold, modifying $g_{\mu\nu}$ and its derivates $R_{\mu\nu}$ and R. The metric tensor affects the geodesics, as we can see in Eq.(1.7), thus a test particle's trajectory in absence of external forces. In practice, what is usually called *gravitational force* becomes the geometry of the spacetime manifold itself, which is determined by how matter and energy are distributed on it ($T_{\mu\nu}$), and determines the trajectories of the bodies on it by means of its own structure.

1.1.1 Cosmology through General Relativity

Cosmology is that field of Physics that studies the Universe at its largest scale. During the last century this sentence completely changed its meaning: we went from believing that our galaxy was the only one in the whole cosmos to knowing that the Universe contains trillions of galaxies and not only it is expanding, but it is also accelerating its expansion. The first major milestone in the history of cosmology was Edwin Hubble's discovery in 1929: all the galaxies are receding from us with a velocity which is proportional to their distance. This can be summarized in the Hubble-Lemaître law:

$$v = H_0 d, \tag{1.15}$$

where v is the recession velocity, d is the distance of the galaxy and H_0 is the Hubble constant. This discovery forced the scientists of the time to embrace expanding models and abandon the *static Universe*, of which also Einstein was a firm believer, as the best and most logical explanation for all the galaxies to recede from us is that the spacetime manifold is in fact expanding, and *dragging* the matter along with itself. Modern cosmology is founded on the assumption that the Universe at large scales is homogeneous and isotropic: this is called *cosmological principle* or *assumption of mediocrity*, since it states the ordinariness of our position in the Universe. In terms of GR, this means that the spacetime manifold can be sliced into foliations of constant time which are perfectly homogeneous and isotropic, thus the Universe at its largest scales is a time-evolving uniform distribution of matter. Assuming the cosmological principle, one can derive the

metric tensor $g_{\mu\nu}$ for a homogeneous and isotropic Universe, the FLRW metric tensor, which using polar comoving coordinates can be written as:

$$g_{\mu\nu} = \begin{pmatrix} -c^2 & 0 & 0 & 0 \\ 0 & \frac{a^2}{1-kr^2} & 0 & 0 \\ 0 & 0 & a^2 r^2 & 0 \\ 0 & 0 & 0 & a^2 r^2 \sin^2(\theta) \end{pmatrix}, \tag{1.16}$$

thus, the line element is:

$$\mathrm{d}s^2 = -c^2 \mathrm{d}t^2 + a^2(t) \left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2 \right) \right], \tag{1.17}$$

where a is the *scale factor* and k is the *curvature constant*.

In this particular set of coordinates, the scale factor multiplies only the spatial part of the metric, thus it can be identified as the factor that encodes information about the expansion of the spatial part of the Universe as a function of the cosmic time. Studying the cosmos at its largest scale, one can consider a as the only degree of freedom of the system *Universe*: simplifying, cosmology is the study of the scale factor as a function of the cosmic time. The curvature constant k determines the geometry of the spatial hypersurfaces of the spacetime manifold and it can be $k = 0, \pm 1$. In the case of $k = 0$, the spatial part of the metric is flat, for $k = 1$ it is a closed 3-sphere and for $k = -1$ it is an open 3-hyperboloid. Its value, as we will see in Section 1.3, depends on the content of the Universe, and observations are constistent with a *spatially flat* Universe, thus with $k = 0$.

Using Eq.(1.17) and defining the proper distance d_{pr} as the distance between two points when $\mathrm{d}t = 0$, by choosing the right reference frame so that $\mathrm{d}\theta = \mathrm{d}\phi = 0$, we can write:

$$d_{pr} = a(t) \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} = a(t)\chi(r), \tag{1.18}$$

where the value of the integral $\chi(r)$ depends on the value of k. In addition to the proper distance, we can define the *comoving distance* d_c: a distance measured at the current time. It only takes a rescaling of the scale factor:

$$d_c = \frac{a_0}{a(t)} d_{pr}(t), \tag{1.19}$$

where a_0 is the scale factor at the current time. For simplicity, we usually set $a_0 = 1$. The proper velocity of a particle can be found by simply deriving Eq.(1.18):

$$v_{pr}(t) = \dot{a}\chi + a\dot{\chi} = v_r + v_{pec}, \tag{1.20}$$

where the second term on the right hand side is the peculiar velocity of the particle v_{pec}, and the first one is what we call the *recession velocity* v_r, meaning the velocity with

which every particle recedes from each other as a result of the expansion of the Universe. As a matter of fact, consistent with what Hubble discovered almost 100 years ago, all space points in the Universe are receding from each other, and their velocity, defining the Hubble parameter as $H \equiv \dot{a}/a$, can be rewritten as:

$$v_r(t) = \dot{a}\chi = H(t)d_{pr}(t). \tag{1.21}$$

In the local Universe (meaning at present time) the Hubble parameter is also called the Hubble constant $H_0 = \dot{a_0}/a_0$, and Eq.(1.21) becomes the Hubble-Lemaître law (Eq.(1.15)). The value of H_0 has been measured several times. For its reference value we will consider to the measurements by Wang et al. (2017) [2], who used the latest *baryonic acoustic oscillations* measurements from the eBOSS survey, finding:

$$H_0 = (67.27 \pm 1.55)\text{km s}^{-1}\text{Mpc}^{-1}. \tag{1.22}$$

Due to the expansion of the Universe the observed light emitted from any kind of cosmic structure appears reddened: all the sources are receding from us, as they are doing for any other observer in the Universe since, as the cosmological principle states, we are not in a privileged reference frame. The light they emit is seen by us with a lower frequency because of their motion with the Hubble flow. This reddening is called *redshift*, z, and it is defined as:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{\Delta\lambda}{\lambda}, \tag{1.23}$$

where λ_{em} is the wavelength of the light emitted by the source, and λ_{obs} is the wavelength of the observed light, after the reddening. Using Eq.(1.17) we can derive a relation between the scale factor at the emission time and at the observation time (a_{em} and a_{obs}) and z, for a light particle moving on null geodesics ($\mathrm{d}s^2 = 0$):

$$\frac{a_{obs}}{a_{em}} = 1 + z, \tag{1.24}$$

which considering an observer at present time, becomes:

$$\frac{a_0}{a(t)} = 1 + z. \tag{1.25}$$

Due to these considerations, the redshift z effectively becomes a quantity with from we can measure distances, since assuming a cosmology (namely $a(t)$) the redshift of a source uniquely depends on its distance from us.

1.2 Content of the Universe: $T_{\mu\nu}$

As stated in Section 1.1, the geometry of the spacetime manifold depends on the right hand side of the field equations ((1.12)): the energy-momentum tensor $T_{\mu\nu}$, which describes the energy and mass distribution on the manifold. Assuming that the content

of the Universe is a perfect fluid, which particles' free path is much smaller than their physical scale of interaction, the energy-momentum tensor can be written as:

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}, \tag{1.26}$$

where ρ is the energy density, p is the pressure and u is the 4-velocity of the fluid element. Since $T^\mu_\nu = g_{\alpha\nu}T^{\mu\nu}$ (from tensorial formalism) and given the expression for $g_{\mu\nu}$ is provided by Eq.(1.16), using units in which $c = \hbar = k_B = 1$, we can derive that:

$$T^\mu_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.27}$$

It can be demonstrated that the covariant derivative of the energy-momentum tensor is null. This provides us the GR version of the continuity and Euler equations:

$$\nabla_\mu T^\mu_\nu = 0. \tag{1.28}$$

These four equations (one for each $\nu = 0, 1, 2, 3$), can be simplified in the following way, assuming isotropy and using Eqs. (1.8) and (1.9) :

$$\frac{\partial \rho}{\partial t} + \frac{\dot{a}}{a}[3\rho + 3p] = 0, \tag{1.29}$$

which, integrating, becomes:

$$\rho_s(a) \propto a^{-3(1+w_s)}. \tag{1.30}$$

In Eq.(1.30), s indicates a generic component of the Universe, with pressure p_s and density ρ_s. We have also used the *equation of state parameter* of the component s, w_s, which is defined as follows:

$$w_s \equiv \frac{p_s}{\rho_s}. \tag{1.31}$$

This parameter is different for every component of the Universe, and in general must be lower than 1, because the sound velocity, defined as:

$$v_s^2 = \left(\frac{\partial p}{\partial \rho}\right)\Bigg|_{S=const}, \tag{1.32}$$

where S is the entropy, must be smaller than c (which in these units is 1). To obtain to Eq.(1.30), we assumed that w_s does not change in time. As we will see in the next Sections though, this assumption micht be relaxed. As hinted before, every component of the Universe has a different value of w_s, specifically:

- **Matter** has $w_m = 0$. In this case Eq.(1.30) reads $\rho_m \propto a^{-3}$, thus indicating that the matter energy density scales like a cubic length, due to the expansion of the volumes in the Universe.
- **Radiation** has $w_r = 1/3$. In this case Eq.(1.30) reads $\rho_m \propto a^{-4}$, and comparing it to the matter energy density evolution, we can see that it fades more quickly as the volumes expand.
- **The cosmological constant** has $w_\Lambda = -1$. In this case equation (1.30) reads $\rho_m \propto$ const, thus the energy density of the cosmological constant, which represents the simplest possible form of *dark energy*, is constant throughout the Universe evolution. The nature of this component will be analyzed in Paragraph 1.4.1.

1.3 The Friedmann equations

In this Section we will work on the Einstein equations in the case of a spatially flat, homogeneous and isotropic Universe, hence assuming a FLRW metric and k = 0. Starting from Eq.(1.12) and setting $g_{\mu\nu}$ as in Eq.(1.16), we can conveniently rewrite the time-time and space-space components of the Ricci tensor from Eq.(1.13), using the definition of the Christoffel symbols (Eq. 1.8):

$$R_{00} = -\delta_{ij}\frac{\partial}{\partial t}\left(\frac{\dot{a}}{a}\right) - \left(\frac{\dot{a}}{a}\right)^2 \delta_{ij}\delta_{ij} = -3\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 = -3\frac{\ddot{a}}{a}, \tag{1.33}$$

$$R_{ij} = \delta_{ij}[2\dot{a}^2 + a\ddot{a}]. \tag{1.34}$$

Now, using Eq.(1.14), we can derive the Ricci scalar:

$$R = g^{\mu\nu}R_{\mu\nu} = g^{00}R_{00} + g^{ij}R_{ij} = -R_{00} + \frac{1}{a^2}R_{ij}, \tag{1.35}$$

which substituting Eqs. (1.33) and (1.34), becomes:

$$R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right]. \tag{1.36}$$

Eventually, if we consider respectively the time-time and space-space components of the Einstein equations:

$$R_{00} - \frac{1}{2}g_{00}R = 8\pi G T_{00}, \tag{1.37}$$

$$R_{ij} - \frac{1}{2}g_{ij}R = 8\pi G T_{ij}, \tag{1.38}$$

and we substitute the expressions for the Ricci tensors the Ricci scalar, the energy-momentum tensor and the metric tensor, we obtain the first and second Friedmann equations, namely:

$$H^2 = \frac{8\pi G}{3}\rho, \tag{1.39}$$

where we used the definition of the Hubble parameter, and:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p). \tag{1.40}$$

If we did not make any assumption on the value of k, the derivation of the first Friedmann equation would have been a bit more complex, and its expression would have become:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.41}$$

which can be rewritten as:

$$\Omega - 1 = \frac{k}{H^2 a^2}, \tag{1.42}$$

where the *density parameter* Ω is defined as follows:

$$\Omega \equiv \frac{8\pi G}{3H^2}\rho = \frac{\rho}{\rho_c}, \tag{1.43}$$

and the *critical energy density* ρ_c, is defined as:

$$\rho_c = \frac{3H^2}{8\pi G}. \tag{1.44}$$

The density parameter is the sum of the individual parameters of every component s of the Universe, as is for the density ρ:

$$\Omega = \sum_s \Omega_s, \tag{1.45}$$

$$\rho = \sum_s \rho_s. \tag{1.46}$$

It is clear from Eq.(1.42) that:

- if $\Omega > 1$ ($\rho > \rho_c$), the right hand side of Eq.(1.42) is positive, thus a rescaling can be performed so that $k = 1$: as seen in Section 1.1.1 the space is *closed*.
- if $\Omega < 1$ ($\rho < \rho_c$), the right hand side of Eq.(1.42) is negative, thus a rescaling can be performed so that $k = -1$: as seen in Section 1.1.1 the space is *open*.

- if $\Omega = 1$ ($\rho = \rho_c$), the right hand side of Eq.(1.42) is null, thus $k = 0$: as seen in Section 1.1.1 the space is *flat*.

Thus, the critical density is that value that if exceeded, it means that the Universe contains enough energy density contrasting the expansion for its geometry to be closed. As we have hinted before, all the studies conducted so far show that the space is approximately *flat*, i.e. k is statistically consistent with 0, thus the total density parameter of our Universe appears consistent with 1. This fact might appear very unlikely: this is called *the flatness problem*, and one of its possible solutions is provided by the *theory of inflation*.

It is interesting to point out that the condition of an *expanding* space depends on the gauge. This is in fact the way we see the Universe if we are a fiducial observer using *comoving* coordinates, which is a very specific point of view. In general, the spacetime manifold is curved by the distribution of matter and energy on it, and if we put ourselves in a comoving reference frame (in which naturally we are not) this curvature appears as an expansion $a(t)$ of the space component: this allows us to treat the Universe in a simil-Newtonian way, even though we should always keep in mind that Newtonian dynamics is not accurate enough at cosmological scales.

1.4 The ΛCDM Model

The Λ-cold dark matter (ΛCDM) model, also called the concordance cosmological model, describes the composition and the evolution of the Universe. It provides a theoretical framework that matches a wide range of cosmological observations, making it a cornerstone of modern cosmology. Estimating the cosmological parameters, such as the ones considered in this thesis, allow us to constrain the model.

The term ΛCDM comes from what are known to be the two main components of the Universe: the cosmologcal constant Λ and the cold dark matter. The first is responsable for the Universe accelerated expansion and the second one refers to non-baryonic and non-relativistic matter, which is not collisional and does not interact electromagnetically, thus cannot be seen in observations. Besides these two main components, there is radiation (γ) which comprises all possible relativistic species, and baryons (b). Each component has its own energy density, thus its density parameter, that contributes to the total density parameter following Eq.(1.45). Generally, CDM and baryons are referred to as *matter* (m), so that $\Omega_m = \Omega_{CDM} + \Omega_b$.

As mentioned before, observations suggest that:

$$\Omega_{tot} = \Omega_\gamma + \Omega_m + \Omega_\Lambda \sim 1, \tag{1.47}$$

and the contributions of each component are [3]:

- $\Omega_\gamma \sim 10^{-5}$

- $\Omega_{CDM} \sim 0.26$
- $\Omega_b \sim 0.05$
- $\Omega_\gamma \sim 0.69$

Assuming these values for the density parameters, we can rewrite (1.30) considering that every component has its own energy density evolution:

$$\rho_s(a) = \Omega_s \rho_c a^{-3(1+w_s)}, \tag{1.48}$$

which shape and normalization depend, respectively, on the equation of state parameter of the component, w_s, and on its density parameter, Ω_s. Inserting Eq.(1.48) for every component in the Friedmann equations, we can derive the evolution of the homogeneous Universe.

1.4.1 Dark energy and the cosmological constant

One of the most discussed components of the ΛCDM model is certainly the *cosmological constant* Λ. It is often described as mysterious, or a *"substance about which we have almost no knowledge at all"* [4]. On the other hand, other physicists have a different opinion, considering the cosmological constant not mysterious at all (see in particular Bianchi & Rovelli, 2010 [5]). As pointed out by Bianchi & Rovelli 2010 [5], talking about a *mysterious substance* to denote dark energy is misleading, as it could be like saying that the centrifugal force that pushes out from a merry-go-round is the effect of a *mysterious substance*. Simply, gravity works in this way: it pushes together at small scales and pulls apart at large scales. In this Paragraph, we summarize what we know and what is yet to be known on the nature of this mysterious (or not!) component of the Universe.

The cosmological constant has been introduced by Einstein in 1917 [6], as an *ad hoc* constant to favour the solutions of a static Universe, rather than an expanding. So, the Einstein Eqs. (1.12) became:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.49}$$

What was born as a repulsive term inserted *ad hoc* by Einstein to make a matter-dominated Universe static, ended up to be one of the pillars of the *standard model of cosmology*. As a matter of fact, the cosmological constant perfectly accurately the accelerated expansion of the Universe, allowing the sum of all the Ω_s terms to be equal to 1 (consistent with the flat, $k = 0$ scenario of observations). Furthermore, this term can be interpreted in two different ways:

- if placed on the left-hand side of Eq.(1.49), it is a repulsive gravity term,i.e. it is part of the geometric structure of the spacetime manifold itself;

- if placed on the right-hand side of Eq.(1.49) it can be seen as a source energy term.

In both cases, it is the simplest form of what is called *dark energy*; it has $w_\Lambda = -1$ and $\rho_\Lambda(a) = \text{const}$, as seen in Section 1.3.

Over the years, the quest for finding out the true nature of dark energy has been addressed by several groups of scientists.

Constraining the dark energy *equation of state parameter* is one of the main goals of the next-generation missions, such as the ESA *Euclid* mission, started in July 2023 [7]. Detecting $w_\Lambda(a) \neq -1$ at any redshift would mean that dark energy is not a constant, but rather a dynamical field [7].

1.5 Growth of structures

As we have seen, the Universe is homogeneous and isotropic if considered at its largest scales. However, if we zoom in and take a closer look at the cosmic web, we find out that at smaller scales the matter distribution is very far from being homogeneous (and we should be thankful for that, otherwise we wouldn't exist). During the first stages of the Universe's life, matter dishomogeneities were slight fluctations of the density field $\rho(x)$, which can conveniently described with the corresponding *overdensity field* $\delta(x)$, defined as follows:

$$\delta(x) = \frac{\rho(x) - \bar{\rho}}{\bar{\rho}}, \tag{1.50}$$

where $\bar{\rho}$ is the mean density of the Universe. The amplitude of these fluctuations at the time of the *last scattering surface* can be derived from the oscillations in temperature observed in the CBM:

$$\frac{\delta T}{T} \sim 10^{-5}. \tag{1.51}$$

The origin of overdensities' distribution is believed to be stochastic, thus their distribution can be described by a Gaussian function centered at zero:

$$p(\delta)\mathrm{d}\delta \propto \exp\left(-\frac{\delta^2}{2\sigma^2}\right), \tag{1.52}$$

where $p(\delta)$ is the probability of having an overdensity δ, and σ is the variance of the Gaussian field. While the Universe kept on expanding, these tiny overdensities, either got erased by the expansion or grew thanks to the attractive nature of gravity at small scales, depending on their size. The Jeans theory provides an analytic explanation of the dynamics at small scales of *linear* density perturbations, and it will be summarized in the next Paragraph.

1.5.1 Jeans theory in an expanding Universe

Let us consider a homogeneous, isotropic and expanding Universe, filled with a perfect fluid with matter density $\rho(x,t)$. Using a simple Newtonian approach, the behaviour of such fluid is described by the following set of equations:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla(\rho \boldsymbol{v}) = 0 \\ \frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\frac{1}{\rho}\nabla p - \nabla\Phi \\ \nabla^2\Phi = 4\pi G\rho \\ p = p(\rho, S) = p(\rho) \\ \frac{\mathrm{d}S}{\mathrm{d}t} = 0 \end{cases}, \tag{1.53}$$

where $\boldsymbol{v}$ is the velocity vector of a fluid element, Φ is the gravitational potential and S is the entropy. The first equation of the system (1.53) is the continuity equation, the second one is the Euler equation, the third one is the Poisson equation, the fourth is the equation of state and the fifth is the adiabatic condition. The latter is assumed since observations suggest that the vast majority of density fluctuation is in fact adiabatic. Now, remembering what we have explained in Paragraph 1.1.1, the velocity of a fluid element is the sum of the Hubble flow and its proper velocity:

$$v = Hr + v_p. \tag{1.54}$$

If we assume to know the exact solutions for the background in an expanding Universe (assuming a null proper velocity of the background), which can be written as:

$$\begin{cases} \rho = \rho_B = \text{const} \\ p = p_B = \text{const} \\ \boldsymbol{v} = Hr \\ \Phi = \Phi_B = \text{const} \end{cases}, \tag{1.55}$$

and we introduce minor perturbations, the perturbed solutions become:

$$\begin{cases} \rho = \rho_B + \delta\rho \\ p = p_B + \delta p \\ \boldsymbol{v} = Hr + \delta\boldsymbol{v} \\ \Phi = \Phi_B + \delta\Phi \end{cases}, \tag{1.56}$$

where $\delta\rho, \delta p, \delta\boldsymbol{v}, \delta\Phi \ll 1$. Neglecting all the terms beyond the linear ones, inserting the solutions of system (1.56) in system (1.53) and converting from the density field $\rho(\boldsymbol{x},t)$ to the overdensity field $\delta(\boldsymbol{x},t)$, if we then transform to Fourier space and search for plane waves as solutions, we obtain the following dispersion relation:

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \delta_k[k^2 v_s^2 - 4\pi G\rho_B] = 0, \tag{1.57}$$

where $k = \frac{2\pi}{\lambda}$ is the wavenumber of a perturbation of scale λ and $v_s \equiv \sqrt{\partial p/\partial \rho}$ is the sound velocity. Solving Eq.(1.57), we can find the time evolution of the overdensities $\delta_k(t)$ for every component of the Universe, and in particular whether they are vanishing, remaining constant or growing, thus favouring the formation of structures. The second term of Eq.(1.57) depends on the assumed cosmology: it is a friction term that grows with $H = \dot{a}/a$, which as we have described is a proxy of the expansion rate of the Universe, that is the large-scale "repulsive nature" of gravity. The higher is the expansion rate, the more the overdensities struggle to grow. Another term that *counters* the formation of structure is the $\delta_k k^2 v_s^2$ one, as it accounts for the characteristic velocity field of the considered fluid. On the other hand, the $-\delta_k 4\pi G \rho_B$ term describes the small-scale "attractive nature" of gravity, which favours the formation of structures.

Considering the evolution of the dark matter densty field, ρ_b, in an Einstein-de Sitter (EdS) cosmology (that is $\Omega_m = 1$, which provides a fairly good approximation in the matter-domination era), substituting v_s with the velocity dispersion (since dark matter is assumed to be non-collisional) and defining the Jeans scale λ_J as:

$$\lambda_J = \frac{2v_s}{5}\sqrt{\frac{6\pi}{G\rho_B}}, \tag{1.58}$$

we obtain the following possible solutions:

- if the scale of the perturbation is $\lambda < \lambda_J$, then the solutions we have $\delta(t) \propto t^{i\alpha}$. In this case the solutions are oscillating and the overdensities are not growing.
- if the scale of the perturbations is $\lambda > \lambda_J$, then the solutions will be the superposition of $\delta_+(t) \propto t^{2/3}$ and $\delta_-(t) \propto t^{-1}$. In the second case the overdensities are reducing, while in the first one they are *growing*.

This means that if the scale of the perturbation is large enough, this one will grow and bring to structure formation.

Dropping the assumption of an EdS Universe, it can be demonstrated that the growing solution becomes:

$$\delta_+(t) = H(t)\int \frac{\mathrm{d}t}{a^2 H^2(t)}. \tag{1.59}$$

This integral does not have an analytic solution. An approximated parametric solution is the following:

$$f \equiv \frac{\mathrm{d}\log\delta_+}{\mathrm{d}\log a} = \Omega_m^\gamma + \frac{\Omega_\Lambda}{70}\left(1 + \frac{1}{2}\Omega_{m,0}\right), \tag{1.60}$$

where f is the linear *growth rate* and $\Omega_{m,0}$ is the value of the matter density parameter at the present day. As Eq.(1.60) shows, the growth rate is strongly dependent on the amount of matter in the Universe:

- if $\Omega_m = 1$ (EdS), then $f = 1$ and $\delta_+ \propto a$,

- if $\Omega_m > 1$, then $f > 1$, thus the density perturbations grow faster than in the EdS,
- if $\Omega_m < 1$, then $f < 1$, thus the density perturbations grow slower thn in the EdS.

It is interesting to point out that f has only a second order dependence on Ω_Λ, meaning that the cosmological constant does not play a fundamental role in the growth of density fluctuations. Furthermore, the value of the exponent of Ω_m, $\gamma = 0.55$, is a direct prediction of GR, meaning that its measurement is of utmost importance in validating GR at cosmological scales: this is also one of the main goals of the *Euclid* space mission [7].

Chapter 2

Clustering

As seen in Section 1.5, the initial distribution of the overdensity field $\delta(\boldsymbol{x})$ is a Gaussian one with null mean (for we are talking about *fluctuations*) and variance σ. Then, the field starts its evolution that leads to the formation of the today's *cosmic web*. For studying the statistical properties of $\delta(\boldsymbol{x})$, it is useful to write it as a sum of waves in Fourier space:

$$\delta(\boldsymbol{x}) = \frac{1}{(2\pi)^3} \int \hat{\delta}(\boldsymbol{k}) \exp(i\boldsymbol{k} \cdot \boldsymbol{x}) \mathrm{d}^3 k, \tag{2.1}$$

while $\delta(\boldsymbol{k})$ can be written by Fourier transforming Eq.(2.1), as:

$$\delta(\boldsymbol{k}) = \int \delta(\boldsymbol{x}) \exp(i\boldsymbol{k} \cdot \boldsymbol{x}) \mathrm{d}^3 x, \tag{2.2}$$

where $\delta(\boldsymbol{x})$ is adimensional, while $\delta(\boldsymbol{k})$ has the dimension of the reciprocal of a volume. It is also useful to define the 3-dimenisonal Dirac delta $\delta_D^{(3)}$ as:

$$\delta_D^{(3)} \equiv \frac{1}{(2\pi)^3} \int \exp(i\boldsymbol{k} \cdot \boldsymbol{x}) \mathrm{d}^3 x, \tag{2.3}$$

which action is:

$$\int F(z') \delta_D^{(3)}(z - z') \mathrm{d}^3 z' = F(z). \tag{2.4}$$

In the next Sections we will analyze the statistical properties of the overdensity field of the Universe. For describing a Gaussian field with null mean we only need the variance, which we will discuss in the next Section. Furthermore, we will see what happens when the distribution becomes *non-Gaussian*.

2.1 Two-point correlation function and power spectrum

One of the most important statistical tools used to investigate the Universe is the *two-point correlation function* (2PCF) $\xi(r)$, defined as follows:

$$\xi(r) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+r)\rangle \,, \tag{2.5}$$

where $\delta(x)$ is the overdensity field at a point $\boldsymbol{x}$ of the Universe and $\delta(x+r)$ is the overdensity field at a point which is at a comoving distance r from $\boldsymbol{x}$. This quantity, related to the variance of the density field, simply self-correlates the overdensity field in two points at a distance r from each other. Passing from configuration space to Fourier space, we can define another quantity which is strictly bonded to the *variance* of the density field, the *power spectrum* (PS) $P(\boldsymbol{k})$:

$$\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k'})\rangle \equiv (2\pi)^3 P(\boldsymbol{k})\delta_D^{(3)}(\boldsymbol{k}+\boldsymbol{k'}). \tag{2.6}$$

A relation between these two quantities can be found substituting Eq.(2.1) into Eq.(2.5):

$$\xi(r) = \frac{1}{(2\pi)^6}\int \mathrm{d}^3k \int d^3k' \,\langle \delta(\boldsymbol{k})\delta(\boldsymbol{k'})\rangle \exp(i\boldsymbol{k}(\boldsymbol{x}+r))\exp(i\boldsymbol{k'}\boldsymbol{x}), \tag{2.7}$$

and sequently, using Eq.(2.6) and the definition of the Dirac delta (Eq. 2.3), we can derive:

$$\xi(r) = \frac{1}{(2\pi)^3}\int \mathrm{d}^3k P(\boldsymbol{k})\exp(i\boldsymbol{k}r), \tag{2.8}$$

that shows that the power spectrum is the Fourier transform of the 2PCF and vice-versa. From Eq.(2.6) we can deduce that:

$$P(\boldsymbol{k}) \propto \langle \delta(\boldsymbol{k})\delta(\boldsymbol{k'})\rangle \propto \left\langle |\delta^2(\boldsymbol{k})|\right\rangle \,, \tag{2.9}$$

where the relation between the power spectrum and the variance of the overdensity field stands out clearly. $P(\boldsymbol{k})$ is actually a *power spectrum density*, since it can be seen from Eq.(2.8) that it is a density in a Fourier volume, and it encodes the information about the significance of perturbations of a given scale $\boldsymbol{k}$ on the total energy distribution: a high value of the PS for high values of $\boldsymbol{k}$ means high energy at small scales, and vice-versa. Moreover, $\boldsymbol{k}$ is three-dimensional. This means that it represents all the combinations of (k_1, k_2, k_3) for which $k = \sqrt{k_1^2+k_2^2+k_3^2}$. Considering the complex conjugate of $\delta(\boldsymbol{k})$, $\delta^*(\boldsymbol{k}) = \delta(-\boldsymbol{k})$, the definition of the power spectrum can be rewritten in the following way, taking $\boldsymbol{k'} = -\boldsymbol{k}$:

$$\langle \delta(\boldsymbol{k})\delta(-\boldsymbol{k})\rangle = \langle \delta(\boldsymbol{k})\delta^*(\boldsymbol{k})\rangle = \left\langle |\delta^2\boldsymbol{k}|\right\rangle = (2\pi)^3 P(\boldsymbol{k})\delta_D^{(3)}(0), \tag{2.10}$$

where $\delta_D^{(3)}(0)$ can be written from Eq.(2.3) as follows:

$$\delta_D^{(3)}(0) = \frac{1}{(2\pi)^3}\int \mathrm{d}^3x = \frac{V_\infty}{(2\pi)^3}, \tag{2.11}$$

where $V_\infty = \int \mathrm{d}^3x$ is the volume of the Universe, which we use as an implicit normalization of δ. Thus, we can write:

$$\left\langle \left|\delta_{\boldsymbol{k}}^2\right| \right\rangle = V_\infty P(\boldsymbol{k}), \tag{2.12}$$

in which the density nature of the power spectrum appears evident. Now, since the overdensity distribution has null mean:

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\delta^2}{2\sigma^2}\right), \tag{2.13}$$

its variance can be written as:

$$\sigma^2 \equiv \left\langle \delta^2(\boldsymbol{x}) \right\rangle. \tag{2.14}$$

By applying the ergodic hypothesis, we can divide the Universe into independent regions large enough to be fair samples, and we can get σ^2 with a double average: the statistical average over all the regions' volumes of the spatial average of the quadratic value of δ in each region, namely:

$$\sigma^2 = \frac{1}{V_\infty}\int_{-\infty}^{\infty} \left\langle \delta^2(\boldsymbol{x}) \right\rangle \mathrm{d}\boldsymbol{x}. \tag{2.15}$$

Now, by applying the Parseval theorem:

$$\int_{-\infty}^{\infty} f(x) g^*(x) \mathrm{d}^n\boldsymbol{x} = \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty} \mathrm{d}k \hat{f}(k)\hat{g}^*(k), \tag{2.16}$$

on Eq.(2.15), we can derive this expression for the punctual variance:

$$\sigma^2 = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} P(k)\mathrm{d}^3k, \tag{2.17}$$

which shows the exact relation between the variance of the overdensity field and the power spectrum.

2.1.1 From the tracers to the density distribution

As we have just seen in the last Paragraph, the matter power spectrum is directly related to the variance of the matter overdensity field. The latter can be inferred from the mass (M) fluctuations in a given comoving volume V:

$$\delta M(\boldsymbol{x}, V) = \frac{M(\boldsymbol{x}, V) - \bar{M}(V)}{\bar{M}(V)}, \tag{2.18}$$

where $\bar{M}(V)$ is the mean mass. This quantity is related to the overdensity field in this way:

$$\delta(\boldsymbol{x}, V) = \frac{\delta M(\boldsymbol{x}, V)}{V}, \tag{2.19}$$

and it is dependent on the considered volume, as this represents the scale with which we are filtering $\rho(\boldsymbol{x})$. The fluctuations of the mass field are computed as the convolution of the overdensity field with a filter function $W(\boldsymbol{x}, R)$, which averages over a given radius R:

$$\delta M(\boldsymbol{x}) = \delta(\boldsymbol{x}) * W(\boldsymbol{x}, R). \tag{2.20}$$

The mass variance σ_M can be obtained in the same way as the *punctual variance*, using $\delta M(\boldsymbol{x})$ instead of $\delta(\boldsymbol{x})$ in Eq.(2.15), which leads to the following relation, since the convolutions become multiplications in Fourier space:

$$\sigma_M^2 = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} P(k) \hat{W}_k^2(k, R) \mathrm{d}^3 k, \tag{2.21}$$

where $\hat{W}_k(k, R)$ is the Fourier transform of the filter function. The mass within a given comoving volume of the Universe can be estimated by counting the *tracers* of the mass distribution contained in it. We can define as *tracers* all visible cosmic structures, such as galaxies, galaxy clusters and cosmic voids. By counting the number of *tracers*, N_t, contained in a given volume of the Universe, we can define the *tracers overdensity field*, $\delta_t(\boldsymbol{x}, V)$, as:

$$\delta_t(\boldsymbol{x}, V) = \frac{N_t(\boldsymbol{x}, V) - \bar{N}_t(V)}{\bar{N}_t(V)}, \tag{2.22}$$

where $\bar{N}_t(V)$ is the mean number of tracers contained in the volume V.

The simplest model to describe the relation between the distribution of the *tracers* and the distribution of the underlying mass, is the following:

$$\delta_t(\boldsymbol{x}, V) = \mathcal{B} \cdot \delta_M(\boldsymbol{x}, V), \tag{2.23}$$

where $\mathcal{B}$ is called the *bias factor*, which relates the overdensities of tracers to those of mass at linear order. The value of $\mathcal{B}$ depends on the mass and redshift of the dark matter halo hosting the selected tracers (Tinker et al., 2010 [8]). While the halo bias can be directly predicted in a given cosmological model, the halo-tracer connection is harder to predict, as it depends on baryonic physics (see for example Katz et al., 1999 [9] and Cen and Ostriker, 2000 [10]). For this reason, the tracers bias parameter must be determined from observations. In the literature several parametrizations to model the bias have been proposed. An extensively used one is the model proposed by Fry & Gaztañaga, 1993 [11], which is a Taylor expansion of δ_g in δ:

$$\delta_t(\boldsymbol{x}) = \sum_{n=0} \frac{b_n}{n!} \delta^n(\boldsymbol{x}). \tag{2.24}$$

The simplest possible approximation can be obtained from Eq.(2.24) by truncating the expansion at the first order:

$$\delta_t(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}), \tag{2.25}$$

where b_1 is the *linear bias parameter*.

We can now define the tracers' power spectrum, $P_t(\boldsymbol{k})$, and the mass power spectrum $P_M(\boldsymbol{k})$ (given Eq. 2.12) by elevating to the power of two both of the members of Eq.(2.23):

$$P_t(\boldsymbol{k}) = \mathcal{B}^2 \cdot P_M(\boldsymbol{k}). \tag{2.26}$$

The higher is the mass, the more correlated are the peaks in the overdensities, and vice-versa: this is the reason why for example the bias factor of galaxy clusters is higher than the one of the galaxies. In the same way, the higher is the redshift, the more correlated are the high-mass structures, thus the bias factor is higher at higher redshifts.

2.1.2 Two-point correlation function of tracers

Let us consider a Universe filled with discrete tracers. In the following, we will call them *galaxies* for simplicity, though the treatment can be generalized to any kind of extra-galactic tracers. If thier mean number density is $\bar{n}$, then the probability of finding one galaxy in an infinitesimal comoving volume $\mathrm{d}V_1$ and ta second galaxy in the infinitesimal comoving volume $\mathrm{d}V_2$ at a distance r from $\mathrm{d}V_1$, if the distribution of the galaxies in the Universe was completely randomic, is:

$$\mathrm{d}P_{12} = \bar{n}^2\mathrm{d}V_1\mathrm{d}V_2. \tag{2.27}$$

However this is *not* the case of our Universe, where instead clustering occurred. The above probability in this case can be written as follows:

$$\mathrm{d}P_{12} = \bar{n}^2\mathrm{d}V_1\mathrm{d}V_2[1 + \xi(r)], \tag{2.28}$$

where $\xi(r)$ is the 2PCF at a distance r between $\mathrm{d}V_1$ and $\mathrm{d}V_2$. In this way, $\xi(r)$ can be seen as the excess or defect of the probability of finding two glaxies at a distance r in our Universe relative to the same probability if the distribution was completely randomic. In particular:

- if $\xi(r) = 0$, the probability of finding two galaxies at a distance r is the same as in a randomic Universe;
- if $\xi(r) > 0$, the probability of finding two galaxies at a distance r is our Universe is higher than in a random distribution. The objects are correlated, and the higher is $\xi(r)$, the more the par separation in $r \pm \delta r$ between pairs of galaxies is likely;
- if $\xi(r) < 0$, the probability of finding two galaxies at a distance r is our Universe is lower than in a random distribution. Thus the objects are anti-correlated.

Furthermore, since the probability must be $\mathrm{d}P_{12} \geq 0$, we deduce that:

$$\xi(r) \geq -1 \tag{2.29}$$

for every separation r. As an example Figure 2.5 shows the 2PCF of luminous red galaxies from the *sloan digital sky survey* (SDSS). As it can be seen, he 2PCF decreases for increasing separations and eventually becomes negative. At intermediate scales, the 2PCF can be described by a negative power law, $\xi(r) \propto r^{-\gamma}$. This has a simple explanation: matter tends to cluster together because of the attractive nature of small-scale gravity, thus is way easier to find a galaxy close to another one than at a large distance from it.

2.2 Three-point correlation function and bispectrum

The standard inflationary theory predicts Gaussian initial conditions for the overdensity field $\delta(\boldsymbol{x})$, thus the primordial matter distribution is fully determined by its first and second moments (i.e. mean and standard deviation). Moreover, the mean of $\delta(\boldsymbol{x})$ is null, since the latter is defined as an *overdensity* field (see Eq. 1.50), that means that it is fully determined by its 2PCF (or power spectrum), connected to the variance as shown in Section 2.1. However, during the growth of structurs nonlinear gravitational instabilities induce *non-Gaussian* signatures in the mass distribution [12]. This non-Gaussianity affects the LSS statistical properties, and it can be directly probed through higher-order statistics, starting from the three-point correlation function (3PCF), or its Fourier transform, the bispectrum (BS). Furthermore, higher-order statistics break the degeneracy between the bias and σ_8, a topic that we will be discussed in Paragraph 2.3.1 [13]. These are the reasons why higher-order moments of the mass distribution are crucial for the study of LSS, thus for cosmology.

The expression for n-order correlation functions, or poly-spectra, can be obtained generalizing Eqs. (2.5) and (2.6). Specifically, the n-point correlation function ξ_N is defined as:

$$\begin{aligned}\xi_N(\boldsymbol{x}_1, ..., \boldsymbol{x}_N) = &\langle \delta(\boldsymbol{x}_1)...\delta(\boldsymbol{x}_N)\rangle_c \langle \delta(\boldsymbol{x}_1)...\delta(\boldsymbol{x}_N)\rangle + \\ &- \sum_{S\in P(\{\boldsymbol{x}_1,...,\boldsymbol{x}_N\})} \prod_{s_i\in S} \xi_{\#s_i}(\boldsymbol{x}_{s_i(1)}, ..., \boldsymbol{x}_{s_i(\#s_i)}).\end{aligned} \tag{2.30}$$

Here the subscript c stands for *connected*, and means that all the connected moments of possible partitions (i.e. ways of dividing a set into nonoverlapping, non-empty subsets such that every element in the set is included in exactly one subset) have been subtracted from the moment, thus it depends only on N-point statistics. The sum is made over the proper partitions $P(\{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\})$ of $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ (namely every partition but the set

itself), and s_i is a subset of $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ contained in the partition S. Moving to Fourier space, the poly-spectra P_N is defined as:

$$\langle \delta(\boldsymbol{k}_1)...\delta(\boldsymbol{k}_N)\rangle_c = (2\pi)^N P_N(\boldsymbol{k}_1, ..., \boldsymbol{k}_N)\delta_D^{(N)}(\boldsymbol{k}_1 + ... + \boldsymbol{k}_N). \tag{2.31}$$

The most simple statistics beyond second order is the three-point statistics, thus the case of $N = 3$. The *connected* 3PCF, $\zeta(r_{12}, r_{13}, r_{23})$, is defined as:

$$\zeta(r_{12}, r_{13}, r_{23}) \equiv \langle \delta(\boldsymbol{r}_1)\delta(\boldsymbol{r}_2)\delta(\boldsymbol{r}_2)\rangle \,, \tag{2.32}$$

where $r_{12} = |\boldsymbol{r}_1 - \boldsymbol{r}_2|$, $r_{13} = |\boldsymbol{r}_1 - \boldsymbol{r}_3|$ and $r_{23} = |\boldsymbol{r}_2 - \boldsymbol{r}_3|$. Similarly the BS, $B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$, is:

$$\langle \delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(\boldsymbol{k}_3)\rangle_c = (2\pi)^3 B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\delta_D^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3). \tag{2.33}$$

The BS is the Fourier transform of the 3PCF and vice-versa, just as the PS and the 2PCF. Following the same arguments as in Paragraph 2.1.2, the probability of finding three galaxies within three comoving volumes elements $\mathrm{d}V_1$, $\mathrm{d}V_2$ and $\mathrm{d}V_3$, centered at the positions $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ is:

$$\mathrm{d}P_{123} = \bar{n}^3 \mathrm{d}V_1 \mathrm{d}V_2 \mathrm{d}V_3 [1 + \xi_3(r_{12}, r_{13}, r_{23})], \tag{2.34}$$

where $\xi_3(r_{12}, r_{13}, r_{23})$ represents the excess or defect of the probability of finding three galaxies in a given triangular configuration $(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3)$ in our Universe (with *clustering*) relative to the one of finding them in the same triangular configuration in a completely randomic Universe. The function ξ_3 is not only influenced by three-point statistics, since the probability of finding a given triangular configuration also depends on the probability of finding two galaxies at a given distance from each other (which is the 2PCF). Indeed we have the following:

$$\xi_3(r_{12}, r_{13}, r_{23}) = \xi(r_{12}) + \xi(r_{13}) + \xi(r_{23}) + \zeta(r_{12}, r_{13}, r_{23}), \tag{2.35}$$

where $\xi(r_{ij})$ is the2PCF of galaxies at positions $r_{12} = |r_1 - r_2|$. The term in equation (2.35) that depends on three-point statistics is $\zeta(r_{12}, r_{13}, r_{23})$, the connected 3PCF. This can be better visualized in Figure 2.1: ξ_3 (in Fourier space in the figure) is the sum of the singlets component (which is null since the mean of $\delta(\boldsymbol{x})$ is null), the pairs components (the three 2PCF of the triangle's sides) and the triplet term, which is the connected 3PCF.

Furthermore, it can be demonstrated that in hierarchical scenarios (in which structures formed through the constant merging of smaller structures) the connected 3PCF, ζ, is proportional to ξ^2 [15]. Thus we can define a new quantity called the *reduced 3PCF*, Q [16], as follows:

$$Q(r_{12}, r_{13}, r_{23}) \equiv \frac{\zeta(r_{12}, r_{13}, r_{23})}{\xi(r_{12})\xi(r_{13}) + \xi(r_{12})\xi(r_{23}) + \xi(r_{23})\xi(r_{13})}, \tag{2.36}$$

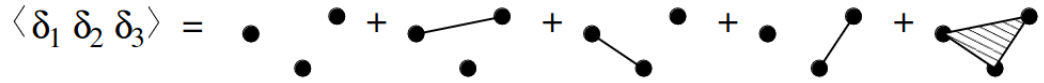


Figure 2.1: *Graphic representation of* $\xi_3(r_{12}, r_{13}, r_{23})$*. The first term is the singlets component (null), then we have the pair components* $\xi(r_{12})+\xi(r_{13})+\xi(r_{23})$*. Eventually there is the connected 3PCF* $\zeta(r_{12}, r_{13}, r_{23})$*. Image of Bernardeau et al., 2002 [14].*

which in some cases can be easier to deal with than the connected 3PCF, as it is characterized by a smaller variation as a function of scales, and its modelling does not depend on the σ_8 parameter, but only on the bias [17].

The modelling of the 3PCF is harder relative to the 2PCF case. Moreover, while the 2PCF is easily representable (see e.g. Figure 2.5), since it depends only on comoving distances between objects pairs, the 3PCF depends on all triangular configurations, that can be also parametrized in different ways. The most common one consists of fixing the length of two sides of the triangles (thus the distance between two tracers) and computing the 3PCF as a function of the third side's length (e.g. Jing et al., 1995 [18], Gaztañaga & Scoccimarro, 2005 [19]). A different choice, proposed by Marín, 2011 [20], consists of fixing two sides of the triangles as follows:

$$\begin{aligned} &r_{12} \\ &r_{13} = u \cdot r_{12}, \end{aligned} \tag{2.37}$$

where u is a constant factor. The third one can be expressed as a function of the angle, θ, between the first two sides:

$$r_{23} \equiv r_{12} \cdot \sqrt{1 + u^2 - 2 \cdot u \cdot \cos\theta}. \tag{2.38}$$

Another way of parametrizing the 3PCF is by fixing all the sides of the triangles (usually in an equilateral configuration) and varying the scales with the same proportions between the sides (e.g. Wang et al., 2004 [21]).

The 3PCF with fixed sides and free angle parametrization is displayed in Figure 2.2. The signal is higher at angles around $0, \pi$, and has a minimum for angles around $\pi/2$. This means that given the position of two galaxies it is more probable to find a third one along the line that connects them. This is because matter tends to distribute in filaments.

The 3PCF is a relatively new statistical tool in clustering, mainly because its precise determination requires a large sample of tracers (see Section 3.3). However, the technological development of the last years allowed researchers to measure the 3PCF and exploit all the benefits that come from its modelling, such as probing primordial non-Gaussianities and breaking degeneracies between cosmological parameters. In 2009

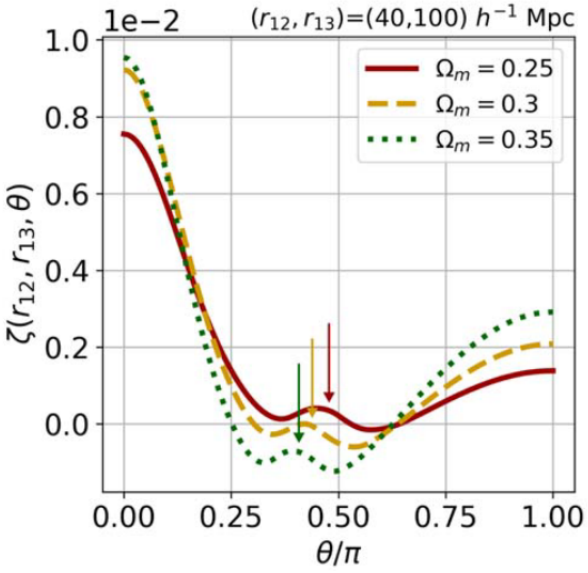


Figure 2.2: *The connected 3PCF in the fixed sides* ($r_{12}, r_{13} = 40$ *and* $100h^{-1}$Mpc*) and free angle* (θ) *parametrization. The continuous red, dashed yellow and dotted green lines represent models with* $\Omega_m = 0.25, 0.3, 0.35$*, respectively. The arrows indicate the BAO peaks. Image from Moresco et al. 2021 [17].*

the first BAO peak detection in the 3PCF has been recorded (Gaztañaga et al., 2009 [22]) in galaxy clustering, while in 2021 it was detected for the first time also in cluster clustering (Moresco et al., 2021 [17]).

2.3 Clustering distortions

Direct measurement of the tracer's density field can be prformed with *redshift surveys*, that allow us to reconstruct the 3D maps of the tracer samples. From the latter it is possible to measure the n-point correlation functions and polyspectra of the tracers. In fact, these measured statistics are not the matter ones, because of the tracer bias (Eq. 2.26). Morover, the tracer positions cannot be inferred directly from the redshifts. This is due to three effects: geometric distortions, dynamic distortions and systematic uncertainties on redshift measurements. The former, i.e. geometric distortions, can occur because of a possibly wrong assumption of the fiducial cosmology when converting the redshifts into comoving distances. On the other hand, neglecting the peculiar velocities of the galaxies which affect the *redshift* coordinate via the Doppler effect introduce a different kind of clustering distortions, called dynamic, or redshift space distortions (RSD). Finally, the tracer maps can be distorted due to systematic redshift errors, which we will refer to in this thesis as *redshift interlopers*. The observed positions of galaxies

are said to be in *redshift space*, while their intrinsic positions are in *real space*. In this Section we will summarize the main properties of these clustering distortions and the way they can be exploited to infer cosmological constraints. The following treatment is mainly based on *Modern Cosmology* by Dodelson & Schmidt [23].

Let us consider a galaxy survey, which maps galaxy spatial distributions by measuring their observed positions:

$$\boldsymbol{x}_{obs}(z, \theta, \phi) = \chi(z)\hat{n}(\theta, \phi), \tag{2.39}$$

where $\hat{n}(\theta, \phi) = \boldsymbol{x}_{obs}/|\boldsymbol{x}_{obs}|$ is the unit vector which identifies the 2D position of the galaxy in the sky plane, and $\chi(z)$ is the comoving distance of the tracer (Eq. 1.19). The latter is retireved by measuring the redshift of the galaxy and assuming a cosmology, thus a scale factor, $a(t)$, and a relation between z and χ, as described in Section 1.1.1. In an actual survey, the fiducial cosmology is not necessarely the unknown true one. We can parametrize the difference with the following relation:

$$\chi_{fid}(z) = \chi(z) + \delta\chi(z). \tag{2.40}$$

Furthermore, the measured redshift of the galaxies can be written as follows, seglecting redshift uncertainties and further subdominant relativistic effects:

$$1 + z = \frac{1}{a_{em}}[1 + u_{\parallel}], \tag{2.41}$$

which is a generalization of Eq.(1.24) by adding the peculiar galaxy velocity along the line of sight, $u_{\parallel} = \boldsymbol{u}_g \cdot \hat{n}$ (and setting $a_0 = 1$). Assuming that the galaxy peculiar velocities are equal, on average, to the matter ones ($\boldsymbol{u}_g = \boldsymbol{u}_m$), which is accurate on sufficiently large scales (see Section 12.6 of [23]), and neglecting second order effects, we can express the total error on the positions as:

$$\Delta\boldsymbol{x} = \delta\chi(z) + \frac{1}{aH}u_{\parallel}(\boldsymbol{x}), \tag{2.42}$$

where the first term on the right hand side is due to the assumption of a wrong fiducial cosmology, while the second term is due to the peculiar velocities of the galaxy along the line of sight. Eq.(2.42) yields:

$$\boldsymbol{x}_{obs} = \boldsymbol{x} + \left(\delta\chi(z) + \frac{1}{aH}u_{\parallel}(\boldsymbol{x})\right)\hat{n}, \tag{2.43}$$

where the total position error is multiplied by $\hat{n}$ since it affects only the component parallel to the line of sight, while the sky plane components are not modified. While these distortions make it more difficult to infer the matter distribution from the observed tracer one, on the other hand they contain plenty of cosmological information, that we will now explain how to extract.

Let us consider an observation in which a certain number of galaxies, with number density $n_{g,obs}$, is detected in a givern volume, V_{obs}. Given Eq.(2.42), so the observed volume and number density are different from the real ones, V and n_g. Since the number of galaxies is the same in real space and redshift space, we have:

$$n_{g,obs}(\boldsymbol{x}_{obs})\mathrm{d}^3x_{obs} = n_g(\boldsymbol{x})\mathrm{d}^3x. \tag{2.44}$$

The comoving 3D volume elements are:

$$\begin{aligned} \mathrm{d}^3x &= x^2\mathrm{d}x\mathrm{d}\Omega, \\ \mathrm{d}^3x_{obs} &= x_{obs}^2\mathrm{d}x_{obs}\mathrm{d}\Omega, \end{aligned} \tag{2.45}$$

where $\mathrm{d}\Omega$ is the angular volume element. Since the latter are the same for the real and the observed volumes, we can write Eq.(**??**) as:

$$n_{g,obs}(\boldsymbol{x}_{obs}) = n_g(\boldsymbol{x})J, \tag{2.46}$$

where the Jacobian, J, is defined as follows:

$$J \equiv \left|\frac{\mathrm{d}^3x}{\mathrm{d}^3x_{obs}}\right| = \left|\frac{\mathrm{d}x}{\mathrm{d}x_{obs}}\right|\frac{x^2}{x_{obs}^2}. \tag{2.47}$$

In the following, we will discard all the nonlinear terms. Assuming that the galaxies are all in a fairly narrow redshift slice around $\bar{z}$, and defining $\delta H(z) = H(z) - H_{fid}(z)$, we can demonstrate that (see Paragraph 11.1.1 of [23]):

$$J \sim \bar{J}\left(1 - \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\right), \tag{2.48}$$

where the constant $\bar{J}$ is:

$$\bar{J} = 1 - 2\frac{\delta\chi(\bar{z})}{\bar{\chi}} + H^{-1}(\bar{z})\delta H(\bar{z}). \tag{2.49}$$

Knowing that the number densities in real and redshift space coordinates are:

$$\begin{aligned} n_g &= \bar{n}_g(1+\delta_g), \\ n_{g,obs} &= \bar{n}_g(1+\delta_{g,obs}), \end{aligned} \tag{2.50}$$

and given Eq.(2.46), we can write:

$$1 + \delta_{g,obs}(\boldsymbol{x}_{obs}) = \bar{J}\left[1 + \delta_g(\boldsymbol{x}[\boldsymbol{x}_{obs}] - \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel(\boldsymbol{x}[\boldsymbol{x}_{obs}])\right]. \tag{2.51}$$

2.3.1 Dynamic distortions

We now have all the ingredients to quanitify the effects of RSD on the power spectrum (thus on the 2PCF). Tracers' velocities are not random, but depend on the matter density field itself. From RSD we can measure the rate at which structure grows at linear order f (Eq. 1.60). The effect is shown in Figure 2.3, where the large-scale scenario is represented in the left panel, while the nonlinear effect at small scales is represented in the right panel. At large (*linear*) scales, the overdense region appears squashed in redshift space, since the matter (thus the tracers) are gravitationally attracted by the center of the region, and thus moving towards it. Hence, the galaxies closer to us move away from us (enhancing the observed redshift) while the galaxies further away from us move towards us (reducing the observed redshift): circular overdensities in real space become elliptical in redshift space, and due to this squashig we expect the clustering to be stronger. At small (*nonlinear*) scales, the displacements of tracer pairs are tipically larger than the distance separating them. As shown in the right panel of Figure 2.3, the effect is an elongation along the line of sight of the contour of constant density, together with a swap of the true and observed positions.

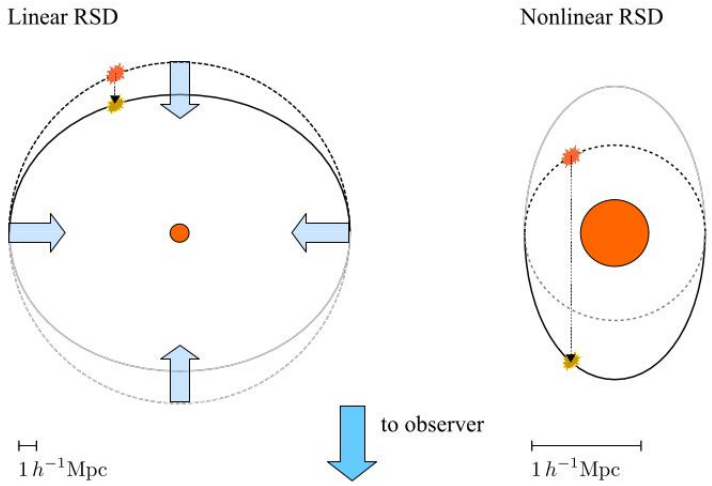


Figure 2.3: *The effect of linear (left panel) and nonlinear (right panel) RSD on the contours of constant density. The contours in real space are indicated with dashed lines, while the ones in redshift space, thus the observed ones, are indicated with continuous lines. Image of Modern Cosmology, by Dodelson & Schmidt [23].*

We consider here only *linear* RSD. In Fourier, space the matter peculiar velocity is related to the *growth factor* by the Boltzmann equation [23]:

$$\boldsymbol{u}_m(\boldsymbol{k},\eta) = aHf\frac{i\boldsymbol{k}}{k^2}\delta(\boldsymbol{k},\eta), \tag{2.52}$$

where η is the conformal time. Assuming the *linear bias relation* given by Eq.(2.25) , in the distant-observer approximation and considering Eq.(2.51) with $\bar{J} = 1$ (thus neglecting for mow the effects of a wrong assumed cosmology), it can be demonstrated that:

$$\delta_{g,RSD}(\boldsymbol{k}) = b_1\delta_m(\boldsymbol{k}) - if\int \mathrm{d}^3x \exp(-i\boldsymbol{k}\cdot\boldsymbol{x})\frac{\partial}{\partial x}\left[\int \frac{\mathrm{d}^3k'}{(2\pi)^3}\exp(i\boldsymbol{k}'\cdot\boldsymbol{x})\delta_m(\boldsymbol{k}')\frac{\boldsymbol{k}'}{k'^2}\cdot\hat{e}_z\right], \tag{2.53}$$

which, by integration, becomes:

$$\delta_{g,RSD}(\boldsymbol{k}) = [b_1 + f\mu_k^2]\delta_m(\boldsymbol{k}), \tag{2.54}$$

where $\mu_k = \hat{e}_z \cdot \hat{k}$, that is the cosine of the angle between the line of sight and the wavevector $\hat{k}$. Eq.(2.54) quantifies what is shown in the left panel of Figure 2.3: RSD cause an enhancement of the overdensities (since $f\mu_k^2$ is always positive) and exists only for perturbations with $\boldsymbol{k}$ parallel to the line of sight (thus with $\mu_k \neq 0$). In *real space*, the overdensities would simply be $\delta_g(\boldsymbol{k}) = b_1\delta_m(\boldsymbol{k})$. Eq.(2.54) can be translated into a relation between the real space and redshift space power spectra:

$$P_{g,RSD}(k,\mu_k,\bar{z}) = P_m(k,\bar{z})[b_1 + f\mu_k^2]^2 + P_N, \tag{2.55}$$

where P_N is a noise term due to the discrete nature of the tracers, which is assumed to be constant. The measured power spectrum $P_{g,RSD}(k,\mu_k,\bar{z})$ varies both with k and μ_k, and this allows us to disentangle the two contributions of the linear bias, b_1, and the linear growth rate, f, that is usually done by performing a multipole decomposition with respect to μ_k. This means that both the galaxy linear bias and the growth factor could be measured. However, we do not know the exact value of the amplitude of the matter power spectrum $P_m(k,\bar{z})$, which is generally parametrized as the amplitude of the linear power spectrum on the scale of $8h^{-1}$Mpc. Thus, by measuring the galaxy power spectrum we can in fact constrain the two parameter combinations $f\,\sigma_8$ and $b_1\,\sigma_8$. As we will see, the degeneracy between f (and b_1) and σ_8 can be broken by using higher-order statistics.

2.3.2 Geometric distortions and baryonic acoustic oscillations

Let us now describe the effects of assuming a wrong cosmology when converting redshift into comoving coordinates, the so-called *Alcock and Paczynski* (AP), or geometric distortiona. As we have mentioned before, we do not know the exact value of the scale factor at all times, $a(t)$, thus the distance-redshift relation that we use to assign the coordinates along the line of sight is affected by a systematic, dependent on the cosmology. The effects of this error are displayed in Figure 2.4. The displacement of the tracers by an amount $\delta\chi(z)$ is, at first order, a rigid translation. However, the difference between

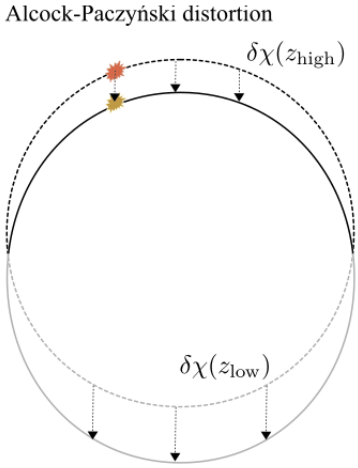


Figure 2.4: *The effect of the assumption of a wrong cosmology (AP distortion) on the contours of constant density. The contours in real space are indicated with a dashed line, while the ones in redshift space, thus the observed ones, are indicated with a continuous line. Image of Modern Cosmology, by Dodelson & Schmidt [23].*

the true and assumed expansion history is actually redshift dependent, thus a circular contour of constant density in real space, becomes an ellipse in redshift space.

Using the flat-sky approximation and calling $\boldsymbol{\theta}$ the 2D position coordinate on the sphere, we can choose the origin in a redshift slice around $\bar{z}$ such that:

$$\boldsymbol{x}_{obs} = 0 \iff \boldsymbol{\theta} = 0, z = \bar{z}. \tag{2.56}$$

The observed coordinates can thus be expressed in this way:

$$\begin{aligned}(x^1_{obs}, x^2_{obs}) &= \chi_{fid}(z) \times (\theta^1, \theta^2), \\ x^3_{obs}(z) &= \chi_{fid}(z) - \chi_{fid}(\bar{z}) \sim \frac{1}{H_{fid}(\bar{z})}(z - \bar{z}),\end{aligned} \tag{2.57}$$

where (x^1_{obs}, x^2_{obs}) are the transverse coordinates and x^3_{obs} is the line-of-sight coordinate of the assigned position. In Eq.(2.57) we have used the linear order expansion in $(z - \bar{z})$ under the assumption of a narrow redshift slice and $\mathrm{d}\chi/\mathrm{d}z = 1/H$. On the other hand, the true transverse coordinates of the objects are:

$$(x^1, x^2) = \chi(z) \times (\theta^1, \theta^2), \tag{2.58}$$

which combined with Eq.(2.57) becomes:

$$(x^1, x^2) = \left[1 - \frac{\delta\chi(z)}{\chi_{fid}(z)}\right](x^1_{obs}, x^2_{obs}). \tag{2.59}$$

Similarly, the true line-of-sight component is:

$$x^3(z) \sim \frac{1}{H(\bar{z})}(z - \bar{z}) = \frac{H_{fid}(\bar{z})}{H(\bar{z})} x^3_{obs}, \tag{2.60}$$

which, defining $\delta H(z) = H(z) - H_{fid}(z)$, becomes:

$$x^3(z) = \left[1 - \frac{\delta H(\bar{z})}{H(\bar{z})}\right] x^3_{obs}. \tag{2.61}$$

If we define $\alpha_\perp$ and $\alpha_\parallel$ as:

$$\begin{aligned} \alpha_\perp &= \left.\frac{\delta\chi}{\chi_{fid}}\right|_{\bar{z}}, \\ \alpha_\parallel &= \left.\frac{\delta H}{H_{fid}}\right|_{\bar{z}}, \end{aligned} \tag{2.62}$$

the relation between the true and observed coordinates can be written as:

$$\boldsymbol{x}(\boldsymbol{x}_{obs}) = \left([1 - \alpha_\perp]x^1_{obs}, [1 - \alpha_\perp]x^2_{obs}, [1 - \alpha_\parallel]x^3_{obs}\right). \tag{2.63}$$

The measurement of $\alpha_\perp$ and $\alpha_\parallel$ allows us to infer, respectively, the comoving distance and the Hubble rate, since Eq.(2.62) can be rewritten as:

$$\begin{aligned} \chi(\bar{z}) &= \chi_{fid}(\bar{z})[1 + \alpha_\perp], \\ H(\bar{z}) &= H_{fid}(\bar{z})[1 + \alpha_\parallel]. \end{aligned} \tag{2.64}$$

Now, the question is: how can $\alpha_\perp$ and $\alpha_\parallel$ be extracted from the measured galaxy power spectrum $P_{g,obs}$? Let us start again from Eq.(2.53), from which we have already computed the effect of RSD. This time we include the factor $\bar{J}$. Defining in the Fourier space $\boldsymbol{k}[\boldsymbol{k}_{obs}]$ as:

$$\boldsymbol{k}[\boldsymbol{x}_{obs}] = \left([1 + \alpha_\perp]k^1_{obs}, [1 + \alpha_\perp]k^2_{obs}, [1 + \alpha_\parallel]k^3_{obs}\right), \tag{2.65}$$

it can be demonstrated that the relation between the observed galaxy and true matter overdensities is encoded in a simple rescaling of the wavevector:

$$\delta_{g,obs}(\boldsymbol{k}_{obs}) = [b_1 + f\mu_k^2]\delta_m(\boldsymbol{k})\Big|_{\boldsymbol{k}=\left([1+\alpha_\perp]k^1_{obs},[1+\alpha_\perp]k^2_{obs},[1+\alpha_\parallel]k^3_{obs}\right)}, \tag{2.66}$$

where we have included also the factor due to RSD. It follows that the redshift space power spectrum is:

$$P_{g,obs}(\boldsymbol{k}_{obs}, \bar{z}) = \left(P_m(k, \bar{z})[b_1 + f\mu_k^2]^2\right)\Big|_{\boldsymbol{k}=\left([1+\alpha_\perp]k^1_{obs},[1+\alpha_\perp]k^2_{obs},[1+\alpha_\parallel]k^3_{obs}\right)} + P_N. \tag{2.67}$$

The fact that the observed power spectrum is exactly the same as the true one but for a simple rescaling is due to the fact that the number of galaxies within a given volume is independent of the used coordinates. Eq.(2.67) contains both the effect of the RSD due to peculiar velocities and the one of a wrong cosmology assumption.

The values of $\alpha_\perp$ and $\alpha_\parallel$ can be inferred from the measurements of the power spectrum using the *barionic acoustic oscillations* (BAO), which is an oscillatory modulation in the power spectrum, roughly of the form $\cos(kr_s)$, where r_s is the sound horizon at recombination ($\sim 105h^{-1}$Mpc). This scale was imprinted in the spatial distribution of the baryonic component of the Universe at the time of recombination, and since baryons and dark matter are coupled by gravity it was transferred to the total matter power spectrum at later times. The BAO feature in the 2PCF is displayed in Figure 2.5 and consists of a peak around r_s, denoting a preference of the sound horizon at recombination as a recurrent pair galaxy semparation. The comoving scale $k_s \sim \pi/r_s$ is well determined by the CMB, thus measuring the BAO scale $k_{obs}[k_s]$ in the observed PS allows us to measure the Jacobian between one set of coordinates and the other, and from which $\alpha_\perp$ and $\alpha_\parallel$. This implies that the BAO feature provides a measurement of $\chi(\bar{z})$ and $H(\bar{z})$ through Eq.(2.64).

2.4 Redshift interlopers

Other than RSD and the AP effect, observed redhsifts can be affected by measurement errors e.g. systematic errors, also called *redshift interlopers*. The new generation of LSS surveys, such as the one performed by the *Euclid* mission, is going deeper and wider in order to find more galaxies and collect better statistics. However, for doing this, they will acquire spectra with relatively low signal-to-noise ratio (SNR) with respect to the ones collected in previous spectroscopic surveys, e.g. in the Sloan Digital Sky Survey (SDSS), and this can lead to less accuracy due to systematic errors on the redshifts. In particular, *Euclid* will determine the redshift of the sources using mainly the single emission line $H\alpha$ and. Given the expected SNR we expect a significant number of spurious determinations due to line mismatches, the so-called *redshift interlopers*. As we have seen in Section 1.1.1, the expansion of the Universe causes the emission lines of almost all extragalactic tracers to appear *redshifted*. This redshift grows with the distance of the source, making it a fundamental quantity for determining the distance of the tracers. However, if only one line of a spectrum is available, and if it is mistaken for another, then a wrong distance is inferred.

The result is that a fraction of the galaxies in a given redshift range actually belongs to other redshift ranges (which one depends on the wavelengths of the mistaken lines) and introduces a spurious systematic effect in the clustering statistics. The interlopers can be of various kinds, though we can divide them in two main groups:

- The **line interlopers** are the tracers for which the *target* line is mistaken for

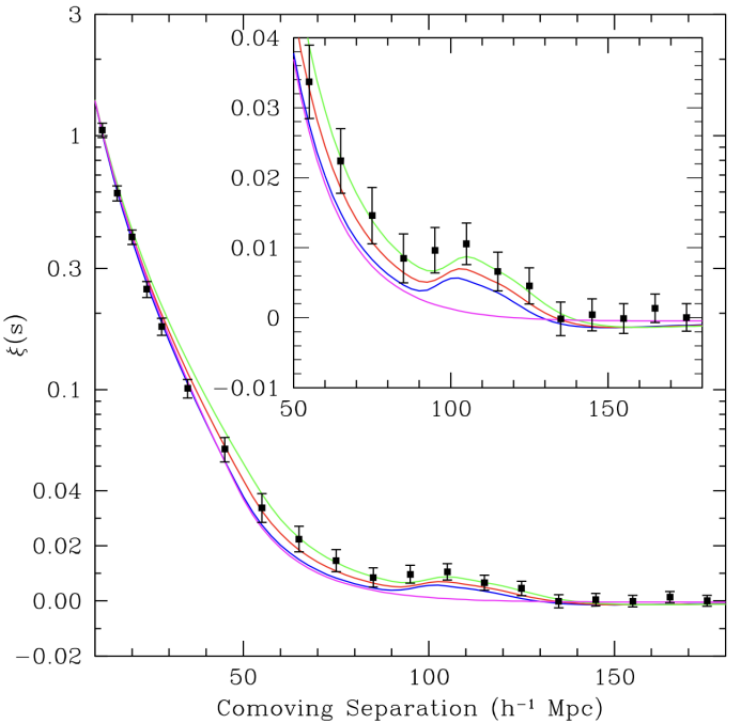


Figure 2.5: *2PCF $\xi(s)$ of a sample of luminous red galaxies (LRGs) from the Sloan Digital Sky Survey (SDSS). The green, red and blue lines represent model predictions with $\Omega_m h^2 = 0.12, 0.13, 0.14$, respectively. The purple line represents a model with $\Omega_b = 0$. The BAO peak around $100h^{-1}$*Mpc *is visible for every model but the last one (since there are no baryons). Image of Eisenstein et al., 2005 [24].*

another emission line. For example in the *Euclid* survey, the target line $H\alpha$ is expected to be possibly mistaken with the $[OIII]$ or $[SIII]$ lines. This kind of interlopers spoils the statistics of a redshift bin, contaminating it with the statistical properties of other cosmological times. This might cause an enhancement in the clustering signal if the interlopers belong to a *lower* redshift bin, or a decrease of the signal if they belong to a *higher* redshift bin.

- The **noise interlopers** are the tracers for which the *target* line is mistaken for a *noise* line: they are very common in surveys which SNR is relatively low. Their clustering signal is a *noise* signal, since there is no relation between the wavelength of the picked line and the target one. Thus their effect is basically a damping of the signal.

The n-point correlation functions and the polyspectra are obviously affected by the presence of interlopers. The study of this effect on both the 2PCF and 3PCF is the

main objective of this work. While RSD and the AP effect are already understood, as we have seen in the previous Sections, the interlopers' effect is a less studied issue, as it is highly accentuated in large and dense surveys with relatively low SNR due to the use of a single emission line for the estimation of the redshift, as in particular in the *Euclid* survey case. This enhances the performance from the point of view of the depth and completeness of the survey, but it has its drawbacks, such as interloper contamination.

The effect of the interlopers on the 2PCF and 3PCF can be dramatic, as they can shift and broaden the BAO peak (see Massara et al. 2021 [25]), other than affecting RSDs. These distortions might strongly bias the constraints on the cosmological parameters. A study conducted by Pullen et al. 2015 [26] using the COSMOS Mock Catalog (CMC), demonstrated that a $0.15 - 0.3\%$ interloper fraction could bias the growth factor by more than 10%. The effect of the redshift interlopers must be studied both on the measurement and the modelling of the correlation functions, in order to understand the way this systematic error affects the cosmological parameter constraints, and if this effect can be mitigated (see Pullen et al. 2015 [26], Addison et al. 2019 [27] and Grasshorn Gebhardst et al. 2019 [28]). The aim of this work is to perform this kind of study, analyzing the effect of the cross-correlations brought by interloper contamination on the computation of both the 2PCF and 3PCF, to understand if the effect on the modelling is a simple rescaling of the bias factor or something that needs more sophisticated modelling.

Chapter 3

Measurement and modelling of the two- and three-point correlation functions

In the last Chapter, we presented the statistical tools we will use for this work, namely the 2PCF, PS, 3PCF, and BS. In this Chapter, we will discuss the application of these tools for our purposes, beginning with the measurement and modeling of the 2PCF and PS, and concluding with the 3PCF and BS. We will review the most common methods used for their measurement and modeling in the literature and introduce new computational methods tailored to our primary objective: understanding the effect of *redshift interlopers* on these statistics.

3.1 Measurement of the two-point correlation function

In Section 2.1, we defined the 2PCF. Let us now discuss the methods for its measurement from surveys, focusing on the approach used in this thesis. As stated in Eq.(2.28), $\xi(r)$ can be seen as the excess or defect of the probability of finding two galaxies at a comoving distance r in our Universe relative to the same probability if the distribution were completely randomic. The measurement of $\xi(r)$ can be performed by counting the pairs of tracers in the survey at a given comoving distance r, referred to as pairs from the *data catalogue*. To estimate the 2PCF, the number of these pairs must be compared with the number of the ones at a comoving distance r in a random sample, known as the *random catalogue*. The latter must match the data catalogue in all selection aspects (i.e. geometry, i.e. sky coverage and depth, and number density of tracers with respect to the redshift, $n(z)$). On the other hand it must not be affected by clustering. Morover, the random sample should be much denser than the data sample (up to 50-60 times), in

order to minimize the Poissonian variance. The ratio between the number of tracers in the random and data catalogues is called the *random-data* ratio, D_R, defined as follows:

$$D_R = \frac{N_R}{N_D}. \tag{3.1}$$

The simplest method to measure the 2PCF from the data-data and random-random pairs, denoted as $DD(r)$ and $RR(r)$ respectively, was proposed by Peebles & Hauser (1974) [29]. Known as the *natural estimator* for its straightforwardness, its expression directly follows from Eq.(2.28), and can be written as:

$$\hat{\xi}_N(r) = \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(r)}{RR(r)} - 1, \tag{3.2}$$

where N_D and N_R denote the number of tracers in the data and random catalogues, respectively. In Eq.(3.2), the prefactor of the ratio of the data-data and random-random pairs, $DD(r)/RR(r)$, is the normalization by the possible combinations of pairs in the entire data and random catalogues. Several other estimators have been proposed in the literature (e.g. Davis & Peebles 1983 [30] and Hamilton 1993 [31]). This work employs the Landy & Szalay (1993) [32] estimator (hereafter LS), which is the most commonly used due to its near-minimal (hence Poissonian) variance [33]. The LS estimator is given by:

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(r)}{RR(r)} - 2\frac{N_R(N_R-1)}{N_D N_R}\frac{DR(r)}{RR(r)}, \tag{3.3}$$

where the data-random cross-pairs $DR(r)$ are also included. These latter are pairs in which one galaxy is from the data catalogue and the other is from the random catalogue.

The 2PCF can be written as follows using a Legendre expansion:

$$\xi(\mu, r) = \sum_l \xi_l(r) P_l(\mu), \tag{3.4}$$

where $P_l(\mu)$ are the Legendre polynomials, μ is the cosine of the angle between the comoving distance of a pair of galaxies, r, and the line of sight. The *multipoles* of the 2PCF, $\xi_l(r)$, are defined as follows:

$$\xi_l(r) = \frac{2l+1}{2}\int_{-1}^{+1} \xi(\mu, r) P_l(r)\,\mathrm{d}\mu. \tag{3.5}$$

By substituting Eq.(3.3) into Eq.(3.5), we obtain the integrated LS estimator of the multipoles of the 2PCF:

$$\hat{\xi}_l^{int}(r) = \frac{2l+1}{2}\int_{-1}^{+1}\left(1 + \frac{N_R(N_R-1)}{N_D(N_D-1)}\frac{DD(\mu,r)}{RR(\mu,r)} - 2\frac{N_R(N_R-1)}{N_D N_R}\frac{DR(\mu,r)}{RR(\mu,r)}\right) P_l(r)\,\mathrm{d}\mu. \tag{3.6}$$

Assuming that $RR(\mu, r) \sim RR(r)$, Eq.(3.6) simplifies to:

$$\hat{\xi}_l^{dir}(r) = \frac{2l+1}{2RR(r)} \left[RR_l(r) + \frac{N_R(N_R-1)}{N_D(N_D-1)} DD_l(r) - 2\frac{N_R(N_R-1)}{N_D N_R} DR_l(r) \right], \quad (3.7)$$

where RR_l, RD_l, and DD_l are the number of pairs weighted by $P_l(\mu)$. Eq.(3.7) provides the expression for the direct LS estimator of the multipoles of the 2PCF, $\hat{\xi}_l^{dir}(r)$, which will be extensively used in this work.

3.1.1 Cross-correlations of the two-point correlation function

In order to study the effect of the redshift interlopers on the 2PCF, it is necessary to disentangle the contributions to the signal coming from the different kinds of tracers: the interlopers and the *pure* galaxies (i.e. those for which the measurement of the redshift is performed using the correct position of the emission lines). This can be done by dividing the total data catalogue into mutually exclusive *sub-catalogues* of different kinds of tracers, each with its own random catalogue, built using the same $n(z)$ as that of the corresponding data. While the method for the construction of such catalogues will be discussed in Chapter 5, here we will focus on the computation and the disentanglement of the signals of the 2PCF from a generic catalogue that can be divided in n mutually exclusive sub-catalogues.

Let us consider a data catalogue, D_{tot}, and its random catalogue, R_{tot}. The former can be divided into n mutually exclusive *data sub-catalogues*, D_i (with $i = 1, ..., n$). To each data sub-catalogue we associate a random sub-catalogue, R_i. The data and random sub-catalogues must be constructed such that:

$$\begin{aligned} D_{tot} &= \sum_i D_i, \\ R_{tot} &= \sum_i R_i. \end{aligned} \quad (3.8)$$

We assume the random-data ratio, D_R, to be the same for each i-th sub-sample. In order to derive an expression that explicits the disentangled signals of the different sub-catalogues, we start from the general expression for an n-point estimator, proposed by Szapudi & Szalay (1993) [34], which can be expressed as:

$$\hat{\xi}^n = \frac{(D-R)^n}{R^n}, \quad (3.9)$$

Using $n = 2$ in Eq.(3.9) the LS estimator for the 2PCF (Eq. 3.3) can be easily derived. In the case of the 2PCF from n sub-catalogues, the signal can be of two kinds: *auto-correlation* and *cross-correlation*. In the following, for brevity, we will refer to the normalized pairs simply as DD, RR, and DR. The *auto-correlations* are the signals that

come from pairs of tracers of the *same* sub-catalogue, and their expression can be easily recovered from Eq.(3.9) as follows:

$$\hat{\xi}_i(r) = \frac{(D_i - R_i)^2}{R_i^2} = \frac{D_i D_i(r) - 2D_i R_i(r) + R_i R_i(r)}{R_i R_i(r)}, \tag{3.10}$$

where D_iD_i, R_iR_i, and D_iR_i are respectively the *normalized auto* data-data, random-random, and data-random pairs of the i-th catalogue. Eqs. (3.10) and (3.3) are exactly the same. Thus, the auto-correlation of the i-th catalogue is merely its 2PCF as if it were the only available data sample. On the other hand, the *cross-correlation* signals are those that come from *mixed* pairs, namely the ones in which one tracer comes from the i-th catalogue and the other a different one, that will be referred to as the j-th catalogue. Their expression can be derived starting from Eq.(3.9), and it reads:

$$\hat{\xi}_{ij}(r) = \frac{(D_i - R_i)(D_j - R_j)}{R_i R_j} = \frac{D_i D_j(r) - D_i R_j(r) - D_j R_i(r) + R_i R_j(r)}{R_i R_j(r)}. \tag{3.11}$$

In the literature, other expressions for the cross-two-point correlation functions are present, such as the one that uses only one random catalogue, R, which is created using the galaxy redshift distribution, $n(z)$, of the combination of the data catalogues. Its expression is the same as the one in Eq.(3.11), using $R_i = R_j = R$, and it is the following:

$$\hat{\xi}_{ij}(r) = \frac{(D_i - R)(D_j - R)}{RR} = \frac{D_i D_j(r) - D_i R(r) - D_j R(r) + RR(r)}{RR(r)}. \tag{3.12}$$

To analyze how the cross and auto-correlation terms contribute to the total signal, we can derive an expression for the *total* 2PCF, starting again from Eq.(3.9):

$$\hat{\xi}_{tot} = \frac{(D_{tot} - R_{tot})^2}{R_{tot}^2}, \tag{3.13}$$

which, substituting with Eq.(3.8), becomes:

$$\hat{\xi}_{tot} = \frac{(\sum_i D_i - \sum_i R_i)^2}{R_{tot} R_{tot}}. \tag{3.14}$$

Performing the algebraic computations and substituting with Eqs. (3.10) and (3.11), we get:

$$\hat{\xi}_{tot}(r) = \sum_i f_i^2 \frac{R_i R_i}{R_{tot} R_{tot}} \hat{\xi}_i(r) + 2 \sum_{i \neq j, i<j} f_i f_j \frac{R_i R_j}{R_{tot} R_{tot}} \hat{\xi}_{ij}(r), \tag{3.15}$$

with $i, j = 1, ..., n$. The factors f_i are defined as follows:

$$f_i = \frac{N_{D_i}}{N_{D_{tot}}}. \tag{3.16}$$

This expression for the total 2PCF estimator has been already used in the last few years (e.g. by Risso et al., in prep. and de la Torre et al., in prep.). It is important to stress the fact that Eq.(3.15) can be derived only by making the assumption of a constant random-data ratio, D_R, for every value of i. Eq.(3.15) provides the general expression for the total 2PCF of a catalogue that can be divided into n mutually exclusive sub-catalogues. Please note that using the expression in Eq.(3.12), it is not possible to disentangle the different contributions coming from the auto and cross-correlation, thus it is not suitable for our purposes. This approach can be generalized to any type of estimator.

3.2 Modelling of the two-point correlation function

In order to constrain cosmological parameters from the measurements of the 2PCF, a model for the PS (its Fourier transform) is generally employed. First, we need to define a model for the *bias*. Using the expression for the bias parameter proposed by Fry & Gaztañaga (1993) [11] (Eq. 2.24) truncated at the second order and using also the *tidal bias parameter*, b_t, the expression for the bias can be written as follows:

$$\delta_g(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) + \frac{b_2}{2}\delta^2(\boldsymbol{x}) + b_t s^2(\boldsymbol{x}), \tag{3.17}$$

where $s^2(\boldsymbol{x})$ is the square of the tidal field. The latter is defined as $s^2(\boldsymbol{x}) = s_{ij}(\boldsymbol{x})s_{ij}(\boldsymbol{x})$, with $s_{ij}(\boldsymbol{x}) = \partial_i\partial_j\Phi(\boldsymbol{x}) - \delta_{ij}\delta(\boldsymbol{x})$. Here, $\Phi(\boldsymbol{x})$ is the gravitational potential, related to the density field by $\nabla^2\Phi(\boldsymbol{x}) = \delta(\boldsymbol{x})$. The tidal bias parameter accounts for nonlocal effects, in particular it encodes information about the influence of large-scale tidal fields on the evolution of matter density. This effect is significant even in the case of local initial conditions (see Catelan et al., 1998 [35]).

Among the several models for the PS proposed in the literature, we utilize the one suggested by Beutler et al. (2014) [36], which is an extended version of the Taruya-Nishimichi-Saito (TNS) model, proposed by Taruya et al. (2010) [37]. Following the notation of Beutler et al., we refer to it as the eTNS model (extended-TNS). The PS of the galaxies in the eTNS model is given by:

$$\begin{aligned} P_g(k,\mu) = & e^{-(fk\mu\sigma_v)^2}\left[P_{g,\delta\delta}(k) + 2f\mu^2 P_{g,\delta\theta}(k) + f^2\mu^4 P_{g,\theta\theta}(k)\right] \\ & + A(k,\mu) + B(k,\mu), \end{aligned} \tag{3.18}$$

where μ is the cosine of the angle between the wavenumber vector, k, and the line of sight, f is the growth rate, σ_v is the one-dimensional velocity dispersion, and θ is the velocity divergence. The first three terms inside the square brackets of Eq.(3.18) are an extension of the Kaiser factor (defined by Kaiser, 1987 [38]), and the A and B terms are corrections due to higher-order correlations between the Kaiser terms and the velocity fields (the definitions can be found in Taruya et al., 2010 [37]), which occur

when mapping to redshift space. The negative exponential factor preceding the brackets is a damping function addressing the fingers-of-God effect. The three power spectra in Eq.(3.18), $P_{g,\delta\delta}$, $P_{g,\theta\theta}$, and $P_{g,\delta\theta}$, are the auto power spectra of density and velocity divergence, and their cross power spectrum, respectively. The distinctive feature of the eTNS model compared to the original TNS model is the inclusion of five galaxy bias parameters. Considering these bias factors, the expressions for the power spectra can be written as follows (Beutler et al., 2014 [36]):

$$\begin{aligned} P_{g,\delta\delta}(k) = & b_1^2 P_{\delta\delta}(k) + 2b_2 b_1 P_{b2,\delta}(k) + 2b_t b_1 P_{bs2,\delta}(k) \\ & + 2n_{3nl} b_1 \sigma_3^2(k) P_{lin}(k) + b_2^2 P_{bss}(k) + 2b_2 b_t P_{b2s2}(k) \\ & + b_t^2 P_{bs22}(k) + N, \end{aligned} \tag{3.19}$$

$$P_{g,\delta\theta} = b_1 P_{\delta\theta}(k) + b_2 P_{b2,\theta}(k) + b_t P_{bs2,\theta}(k) + b_{3nl}\sigma_3^2(k) P(k), \tag{3.20}$$

where $P(k)$ is the linear matter PS, b_1 is the linear bias parameter, b_2 is the nonlinear bias parameter, and b_t and b_{3nl} are the tidal and third-order non-local bias parameters, respectively. The latter are significant for explaining the large-scale PS (Saito et al., 2014 [39]). N is a constant stochasticity term accounting for random or unknown processes that affect the galaxy distribution. The expressions for the other PS present in Eqs. (3.19) and (3.20) can be found in Beutler et al. (2014) [36]. Since Beutler et al. (2014) assumes no velocity bias, the auto PS of velocity divergence, $P_{g,\theta\theta}$, is simply:

$$P_{g,\theta\theta} = P_{\theta\theta}, \tag{3.21}$$

3.3 Measurement of the three-point correlation function

This Section outlines the methods for measuring the 3PCF. For simplicity, we will refer to the $\zeta = \zeta(r_{12}, r_{13}, \theta)$ parametrization of the connected 3PCF described in Section 2.2, where r_{12} and r_{13} are the fixed sides of the triangular configuration, and θ is the angle between them. The concept is similar to the measurement of the 2PCF but this time it involves *triplets* of tracers instead of pairs. As mentioned in Section 2.2, $\zeta(r_{12}, r_{13}, \theta)$ represents the excess or defect of probability of finding three tracers in a given triangular configuration (r_{12}, r_{13}, θ) in our Universe relative to the one of finding them in same configuration if the distribution were completely random.

The classical method for the measurement of the 3PCF involves counting the triplets in the data catalogue and comparing them with those in the random catalogue, analogously to the 2PCF approach. The most commonly used estimator for the 3PCF is the

one proposed by Szapudi & Szalay (1998) (hereafter, SS estimator) [34], which can be expressed as follows:

$$\hat{\zeta}(r_{12}, r_{13}, \theta) = \frac{DDD(r_{12}, r_{13}, \theta) - 3DDR(r_{12}, r_{13}, \theta) + 3DRR(r_{12}, r_{13}, \theta) - RRR(r_{12}, r_{13}, \theta)}{RRR(r_{12}, r_{13}, \theta)}, \tag{3.22}$$

The SS estimator corresponds to Eq.(3.9) for $n = 3$. Here, $DDD(r_{12}, r_{13}, \theta)$, $DDR(r_{12}, r_{13}, \theta)$, $DRR(r_{12}, r_{13}, \theta)$, and $RRR(r_{12}, r_{13}, \theta)$ are the normalized counts of data-data-data, data-data-random, data-random-random, and random-random-random triplets, respectively. For brevity, these will be referred to as DDD, DDR, DRR, and RRR. Their normalizations are the following:

$$\begin{aligned} \text{Norm}_{DDD} &= N_D(N_D - 1)(N_D - 2), \\ \text{Norm}_{DDR} &= N_D(N_D - 1)N_R, \\ \text{Norm}_{DRR} &= N_D N_R(N_R - 1), \\ \text{Norm}_{RRR} &= N_R(N_R - 1)(N_R - 2). \end{aligned} \tag{3.23}$$

Even though searching algorithms such as chain-mesh[1] can speed up the counting, this estimator remains computationally expensive. Counting the triplets requires $O(N^3)$ operations, where N is the number of tracers in the catalogue. This is significantly more than the $O(N^2)$ operations needed for counting pairs in the 2PCF, particularly for random catalogues, which can be up to 50 times denser than the data ones. To address this, Slepian & Eisenstein (2015a) [40] proposed a new method for measuring the 3PCF based on Spherical Harmonic Decomposition (SHD), referred to as SE15.

The classical algorithm for computing the 3PCF scales with the number of possible triangles in a survey. To measure the 3PCF up to a scale $R_{\max}$, the number of possible triangles in the catalog is $N(nV_{R_{\max}})^2$, where N is the number of galaxies, n is the number density, and $V_{R_{\max}}$ is the volume of the sphere with radius $R_{\max}$. SE15 decomposes the connected 3PCF into multipoles, following Szapudi's (2004) [41] proposal:

$$\zeta(r_{12}, r_{13}, \mu) = \sum \zeta_\ell(r_{12}, r_{13}) P_\ell(\mu), \tag{3.24}$$

where $\mu = \cos\theta$ and $P_\ell(\mu)$ are the Legendre Polynomials. The quantity to be computed is now the radially binned ζ_ℓ. Instead of computing the angle between every possible pair of galaxies within $V_{R_{\max}}$ for each tracer, SE15 uses the spherical harmonic addition

[1]The chain-mesh method is used for counting pairs and triplets of objects within a specified range of scales. Initially, the catalogue is divided into cubic cells, with the indexes of all objects in each cell stored in vectors. To locate objects near a given one, the search is limited to the cells within the chosen scale range, reducing the number of unnecessary counts of objects at larger distances. Thus, the method's efficiency mainly depends on the ratio between the scale range of the search area and the maximum separation between objects in the catalogue. This method is implemented in the *CBL* (Marulli et al, 2016 [1]).

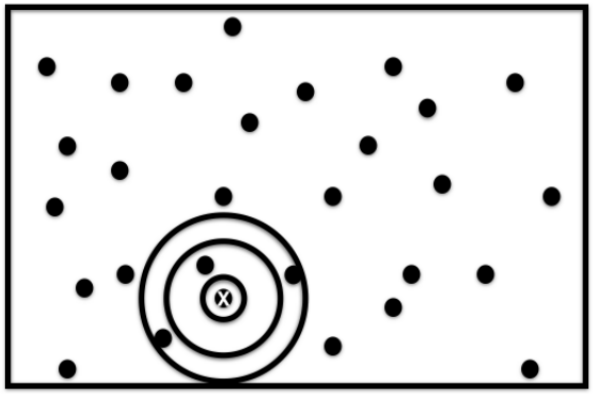


Figure 3.1: *Visualization of the SE15 method. Galaxies are indicated by black dots. The algorithm centers on a galaxy (white cross) and computes the spherical harmonic coefficients $a_{\ell m}$ of δ around it, mapping the tracer overdensity field (black circles). The combination of the $a_{\ell m}$ coefficients for all galaxies yields the connected 3PCF. Image of Slepian & Eisenstein, 2015a [40].*

theorem to decompose the Legendre polynomials into factors dependent on a single angular variable each. It can be shown that the radially binned connected 3PCF can be expressed as follows:

$$\hat{\bar{\zeta}}_\ell(r_{12}, r_{13}, \boldsymbol{s}) = \frac{1}{4\pi}\delta(\boldsymbol{s}) \sum_{m=-\ell}^{\ell} a_{\ell m}(r_{12}, \boldsymbol{s}) a^*_{\ell m}(r_{13}, \boldsymbol{s}), \tag{3.25}$$

where $\boldsymbol{s}$ denotes the position of every possible origin, $\delta(\boldsymbol{s})$ is the overdensity field around the origin, and $a_{\ell m}$ are the SHD coefficients of the tracer overdensity field. The operations for the computation of the latter scale as $nV_{R_{\max}}$, thus the total operations required for computing the multipoles of the connected 3PCF for every galaxy are $N(nV_{R_{\max}})$. This method saves a factor of $nV_{R_{\max}}$ in time, making the computational time comparable to that of the 2PCF. Moreover, since the shape of the connected 3PCF for fixed side lengths as a function of the angle is expected to be smooth, only a few multipoles (~ 10) are needed to recover the angular dependence [40]. A visualization of the algorithm is shown in Figure 3.1.

3.3.1 Cross-correlations of the three-point correlation function

This Paragraph proposes a new expression for the estimator for the 3PCF, based on the SS one, which can disentangle the contributions of different sub-catalogues to the total 3PCF. Each of the n sub-catalogues (as defined by Eq. 3.8) provides galaxies that contribute to the triplet counts used in the computation of the 3PCF. Consequently,

each catalogue will contribute a certain number of *auto-triplets* (triplets of galaxies all from the same catalogue) and *cross-triplets* (triplets consisting of galaxies from different catalogues). The former results in n *auto-signals* from the n different catalogues, while different combinations of the latter generate *cross-signals*. This estimator, similarly to the one described in Paragraph 3.1.1 for the 2PCF, is crucial for studying the effects of redshift interlopers on the measurement and modelling of the 3PCF.

We begin by deriving the expressions for the auto and cross-correlation signals using the general expression for a three-point estimator proposed by Szapudi & Szalay (1993) [34]:

$$\hat{\zeta} = \frac{(D-R)^3}{R^3}. \tag{3.26}$$

For the auto-correlations ζ_{iii}, only the $i-i-i$ type of triplets are considered, so Eq.(3.26) becomes:

$$\hat{\zeta}_{iii} = \frac{(D_i - R_i)^3}{R_i^3} = \frac{D_iD_iD_i - 3D_iD_iR_i + 3D_iR_iR_i - R_iR_iR_i}{R_iR_iR_i}, \tag{3.27}$$

where $D_iD_iD_i$, $D_iD_iR_i$, $D_iR_iR_i$, and $R_iR_iR_i$ are the normalized auto data-data-data, data-data-random, data-random-random, and random-random-random triplet counts of the i-th catalogue, respectively.

Next, we address the cross-correlations. For the 3PCF, there are different types of cross-terms:

- **Bi-cross-terms**: which arise from triplets where the galaxies come from two different sub-catalogues, such as the $i-i-j$ and $i-j-j$ types of triplets. We refer to these as ζ_{iij} and ζ_{ijj}.
- **Tri-cross-terms**: which arise from triplets where the galaxies come from three different sub-catalogues, such as the $i-j-k$ types of triplets. We refer to these as ζ_{ijk}.

For ζ_{iij}, Eq.(3.26) can be written as:

$$\begin{aligned} \hat{\zeta}_{iij} &= \frac{(D_i - R_i)^2(D_j - R_j)}{R_iR_iR_j}, \\ \hat{\zeta}_{iij} &= \frac{D_iD_iD_j - D_iD_iR_j - 2D_iR_iD_j + 2D_iR_iR_j + R_iR_iD_j - R_iR_iR_j}{R_iR_iR_j}. \end{aligned} \tag{3.28}$$

Similarly, for ζ_{ijj} we can write:

$$\begin{aligned} \hat{\zeta}_{ijj} &= \frac{(D_i - R_i)(D_j - R_j)^2}{R_iR_jR_j}, \\ \hat{\zeta}_{ijj} &= \frac{D_iD_jD_j + D_iR_jR_j - 2D_iD_jR_j + 2R_iD_jR_j - R_iD_jD_j - R_iR_jR_j}{R_iR_jR_j}. \end{aligned} \tag{3.29}$$

Finally, the expression for the tri-cross-terms ζ_{ijk} is:

$$\hat{\zeta}_{ijk} = \frac{(D_i - R_i)(D_j - R_j)(D_k - R_k)}{R_i R_j R_k},$$
$$\hat{\zeta}_{ijk} = \frac{D_i D_j D_k - D_i R_j D_k - D_i D_j R_k - R_i D_j D_k + R_i R_j D_k + R_i D_j R_k + D_i R_j R_k - R_i R_j R_k}{R_i R_j R_k}. \tag{3.30}$$

To understand the contribution of each term, we express the estimator of the total three-point correlation function from Eq.(3.26) as follows:

$$\hat{\zeta}_{tot} = \frac{(D_{tot} - R_{tot})^3}{R_{tot}^3}, \tag{3.31}$$

which, substituting with Eq.(3.8), becomes:

$$\hat{\zeta}_{tot} = \frac{(\sum_i D_i - \sum_i R_i)^3}{R_{tot}^3}. \tag{3.32}$$

After performing the algebraic computations and substituting Eqs. (3.27), (3.28), (3.29), and (3.30) we obtain:

$$\begin{aligned}\hat{\zeta}_{tot} = &\sum_i f_i^3 \frac{R_i R_i R_i}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{iii} + 3 \left[\sum_{i \neq j, i<j} \left(f_i^2 f_j \frac{R_i R_i R_j}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{iij} + f_i f_j^2 \frac{R_i R_j R_j}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{ijj} \right) \right] \\ &+ 6 \sum_{i \neq j \neq k, i<j<k} f_i f_j f_k \frac{R_i R_j R_k}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{ijk},\end{aligned} \tag{3.33}$$

where f_i is the ratio of the number of tracers in the i-th sub-catalogue to the total number of tracers (Eq. 3.16). It is important to highlight that Eq.(3.33) can be derived only by making the assumption of a constant random-data ratio, D_R, for every value of i. Eq.(3.33) provides the general expression for the composition of the total connected 3PCF of a catalogue divided into n mutually exclusive sub-catalogues. This approach can be generalized to any type of estimator.

3.4 Modelling of the three point correlation function

Once we measure the 3PCF using the methods described in the previous Section, we must apply a model for the 3PCF (or the BS) to constrain the cosmological parameters. The model used in this work is the one proposed by Alessandro Benati (2023) [42] (hereafter B23), which extends the Scoccimarro et al. (1999) [43] model (hereafter SCF99, following the notation of [42]) for the BS and revises the Slepian & Eisenstein (2017) [44] model (hereafter SE17, following the notation of [42]) for the 3PCF.

3.4.1 Modelling of the bispectrum

As previously mentioned, the B23 model for the BS is an extension of SCF99 that incorporates the tidal bias parameter b_t, whose significance has been discussed in Section 3.2. Following the derivation of SCF99, we define the first- and second-order perturbation theory (PT) redshift-space kernels Z_n for the galaxy density field, supplemented with the B23 model's tidal bias term $b_t S_2(\mathbf{k}_1, \mathbf{k}_2)$:

$$Z_1(\mathbf{k}_i) = (b_1 + f\mu_i^2), \tag{3.34}$$

$$\begin{aligned} Z_2(\mathbf{k}_1, \mathbf{k}_2) = & b_1 F_2(\mathbf{k}_1, \mathbf{k}_2) + f\mu^2 G_2(\mathbf{k}_1, \mathbf{k}_2) + \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2)\right. \\ & \left. + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right] + \frac{b_2}{2} + b_t S_2(\mathbf{k}_1, \mathbf{k}_2), \end{aligned} \tag{3.35}$$

where $\mu \equiv \mathbf{k} \cdot \hat{z}/k$, $\mathbf{k} \equiv \mathbf{k}_1 + \ldots + \mathbf{k}_n$, and $\mu_i \equiv \mathbf{k}_i \cdot \hat{z}/k_i$. The second-order kernel, S_2, for the tidal tensor, s_{ij}, is defined as follows:

$$S_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} - 1 = x^2 - 1, \tag{3.36}$$

where $x \equiv \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \equiv \cos\theta$. The second-order kernels for the real-space density and velocity-divergence fields, F_2 and G_2 respectively, are defined as:

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{x}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2x^2}{7}, \tag{3.37}$$

$$G_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{3}{7} + \frac{x}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{4x^2}{7}. \tag{3.38}$$

We can define the three-level redshift-space BS as follows:

$$B_s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2Z_2(\mathbf{k}_1, \mathbf{k}_2) Z_1(\mathbf{k}_1) Z_1(\mathbf{k}_2) P(k_1) P(k_2) + \text{cyc.}, \tag{3.39}$$

where 'cyc.' denotes cyclic permutations of $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$. Hereafter, we neglect the terms above fourth-order, and we adopt the galaxy at $\mathbf{k}_1$ as the reference frame origin. The latter is assumed to consistently contribute to the second-order density field. This yields the *pre-cyclic* solution, prior to cyclic summation. Upon azimuthal averaging (i.e. over ϕ), B23 decomposes the BS using Legendre polynomials, and substitutes the PT redshift-space kernels, Z_n, into Eq.(3.39), obtaining the following:

$$\begin{aligned} B_s^{(\ell)}(k_1, k_2, \theta) = & P(k_1) P(k_2) b_1^4 \left[F_2(\mathbf{k}_1, \mathbf{k}_2) D_{SQ1}^{(\ell)} + G_2(\mathbf{k}_1, \mathbf{k}_2) D_{SQ2}^{(\ell)}\right. \\ & \left. + (\gamma + 2\gamma_t S_2(\mathbf{k}_1, \mathbf{k}_2))\, D_{NLB}^{(\ell)} + D_{FOG}^{(\ell)}\right] + \text{cyc.}, \end{aligned} \tag{3.40}$$

where $\gamma = b_2/b_1$, $\gamma_t = b_t/b_1$, θ is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$, and:

$$D^{(0)}_{SQ1} = \frac{2(15 + 10\beta + \beta^2 + 2\beta^2 x^2)}{15 b_1}, \tag{3.41}$$

$$\begin{aligned} D^{(0)}_{SQ2} = & 2\beta(35k_1^2 + 28\beta k_2^2 + 3\beta^2 k_1^2 + 35k_2^2 + 28\beta k_1^2 + 3\beta^2 k_2^2 + 70k_1k_2x \\ & + 84\beta k_1 k_2 x + 18\beta^2 k_1 k_2 x + 14\beta k_1^2 x^2 + 12\beta^2 k_1^2 x^2 + 14\beta k_2^2 x^2 \\ & + 12\beta^2 k_2^2 x^2 + 12\beta^2 k_1 k_2 x^3)/(105 k_3^2 b_1), \end{aligned} \tag{3.42}$$

$$D^{(0)}_{NLB} = \frac{(15 + 10\beta + \beta^2 + 2\beta^2 x^2)}{15 b_1}, \tag{3.43}$$

$$\begin{aligned} D^{(0)}_{FOG} = & \beta(210k_1k_2 + 210\beta k_1 k_2 + 54\beta^2 k_1 k_2 + 6\beta^3 k_1 k_2 + 105 k_1^2 x \\ & + 189\beta k_1^2 x + 99\beta^2 k_1^2 x + 15\beta^3 k_1^2 x + 105 k_2^2 x + 189\beta k_2^2 x \\ & + 99\beta^2 k_2^2 x + 15\beta^3 k_2^2 x + 168\beta k_1 k_2 x^2 + 216\beta^2 k_1 k_2 x^2 \\ & + 48\beta^3 k_1 k_2 x^2 + 36\beta^2 k_1^2 x^3 + 20\beta^3 k_1^2 x^3 + 36\beta^2 k_2^2 x^3 \\ & + 20\beta^3 k_2^2 x^3 + 16\beta^3 k_1 k_2 x^4)/(315 k_1 k_2). \end{aligned} \tag{3.44}$$

where $\beta = f/b_1$. $D^{(\ell)}_{SQ1}$ and $D^{(\ell)}_{SQ2}$ account for the first- and second-order contributions to large-scale squashing (Kaiser effect or pancakes-of-God), $D^{(\ell)}_{NLB}$ represents the nonlinear bias contribution, and $D^{(\ell)}_{FOG}$ accounts for the damping effect due to velocity dispersion (fingers-of-God). This analysis considers the monopole term exclusively.

3.4.2 Modelling of the three-point correlation function

The SE17 model for the redshift-space 3PCF is derived by transforming the SCF99 model from Fourier space to configuration space. The B23 update introduces a dependence on the tidal bias and corrects for a linear bias factor, b_1, in the expression of certain BS multipoles (see Benati, 2023 [42] for the detailed discussion). In the following, for simplicity, we will refer to the comoving sides of the triangular configurations (r_{12}, r_{13} and r_{23}) as r_1, r_2 and r_3, respectively. The concept behind SE17 involves leveraging the Legendre expansion of the 3PCF in redshift space:

$$\zeta(r_1, r_2, \hat{r_1} \cdot \hat{r_2}) = \sum_\ell \zeta_\ell(r_1, r_2) P_\ell(\hat{r_1} \cdot \hat{r_2}), \tag{3.45}$$

where P_ℓ are Legendre polynomials and the coefficients ζ_ℓ are given by:

$$\begin{aligned} \zeta_\ell(r_1, r_2) = \frac{2\ell + 1}{2} \int_{-1}^{1} \mathrm{d}x_{12} P_\ell(x_{12}) \left[\zeta_{\mathrm{pc}}(r_1, r_2, x_{12}) + \zeta_{\mathrm{pc}}(r_1, r_3, x_{13})\right. \\ \left. + \zeta_{\mathrm{pc}}(r_2, r_3, x_{23})\right], \end{aligned} \tag{3.46}$$

where x_{ij} represents the cosine of the angle between r_i and r_j, and $\zeta_{\rm pc}$ denotes the pre-cyclic 3PCF. It has been demonstrated (see Benati, 2023 [42] and Slepian & Eisenstein, 2015a [45]) that by expanding the BS monopole using Legendre polynomials, a relation can be established between the multipoles of the pre-cyclic 3PCF $\zeta_{{\rm pc},\ell}$ and the BS multipoles $B_{s,\ell}$. The relation can be written as follows:

$$\zeta_{{\rm pc},\ell}(r_1, r_2) = (-1)^\ell \int {\rm d}k_1 {\rm d}k_2 \frac{k_1^2 k_2^2}{(2\pi^2)^2} B_{s,\ell}(k_1, k_2) j_\ell(k_1 r_1) j_\ell(k_2 r_2), \tag{3.47}$$

where j_ℓ is the spherical Bessel function of order ℓ. By expanding all terms in the BS expression (3.40) into multipole moments and using Eq.(3.47), it has been shown that

the multipole moments of the line-of-sight averaged 3PCF B23 model can be written as:

$$
\begin{aligned}
\ell = 0: \quad & \xi_1^{[0]}\xi_2^{[0]}\Bigg\{ b_1^4\left[\frac{2}{3}\beta + \frac{38}{45}\beta^2 + \frac{2}{5}\beta^3 + \frac{2}{25}\beta^4\right] \\
& + b_1^3\Bigg[\frac{34}{21}\left(1 + \frac{47}{51}\beta + \frac{163}{425}\beta^2 + \frac{201}{2975}\beta^3\right) \\
& + \gamma\left(1 + \frac{2}{3}\beta + \frac{1}{9}\beta^2\right) - \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{7}{75}\beta^2\right)\Bigg]\Bigg\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_0(r_1, r_2) \\
\ell = 1: \quad & - \left[\xi_1^{[1+]}\xi_2^{[1-]} + \xi_2^{[1+]}\xi_1^{[1-]}\right]\Bigg\{ b_1^4\left[\frac{1}{3}\beta + \frac{3}{5}\beta^2 + \frac{67}{175}\beta^3 + \frac{3}{35}\beta^4\right] \\
& + b_1^3\left[1 + \beta + \frac{37}{75}\beta^2 + \frac{17}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_1(r_1, r_2) \\
\ell = 2: \quad & \xi_1^{[2]}\xi_2^{[2]}\Bigg\{ b_1^4\left[\frac{16}{45}\beta^2 + \frac{16}{35}\beta^3 + \frac{32}{245}\beta^4\right] \\
& + b_1^3\Bigg[\frac{8}{21}\left(1 + \frac{4}{3}\beta + \frac{54}{35}\beta^2 + \frac{111}{245}\beta^3\right) \\
& + \frac{4\beta^2\gamma}{45} + \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{1}{21}\beta^2\right)\Bigg]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_2(r_1, r_2) \\
\ell = 3: \quad & - \left[\xi_1^{[3+]}\xi_2^{[3-]} + \xi_2^{[3+]}\xi_1^{[3-]}\right]\Bigg\{ b_1^4\left[\frac{8}{175}\beta^3 + \frac{8}{315}\beta^4\right] \\
& + b_1^3\left[\frac{8}{75}\beta^2 + \frac{8}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_3(r_1, r_2) \\
\ell = 4: \quad & \xi_1^{[4]}\xi_2^{[4]}\Bigg\{ \frac{128}{11025}\beta^4 b_1^4 + b_1^3\left[-\frac{32}{3675}\beta^2 + \frac{32}{8575}\beta^3 + \frac{32}{525}\beta^2\gamma_t\right]\Bigg\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_4(r_1, r_2) \\
\ell \geq 5: \quad & b_1^3\beta^2\left(7 + 3\beta\right)\kappa_\ell(r_1, r_2).
\end{aligned}
\tag{3.48}
$$

Here, $\xi_i^{[n]}$, $\xi_i^{[n\pm]}$, and κ_ℓ are defined in Eqs. (4.37) and (4.39) of Benati (2023) [42].

Chapter 4

New functions for the cross-terms of the two- and three-point correlation functions

All computations in this study were conducted using CBL [1], a comprehensive set of *free software* C++/Python libraries designed for cosmological calculations. These libraries offer a versatile computational framework suitable for the a wide range of cosmological studies, including the computation and modelling of the 2PCF and 3PCF. The CBL initially lacked classes and functions necessary for computing the cross-terms of these correlation functions using the estimators discussed in Sections 3.1.1 and 3.3.1. As demonstrated earlier, these estimators are pivotal for any future investigations of the impact of redshift interlopers on the statistical distribution of matter. Therefore, we have implemented and rigorously tested four new classes and several functions to bridge this gap. This Chapter presents our contributions, beginning with the computation of the classic cross-2PCF, followed by discussions on measuring its first three even multipoles. Eventually, we will present the new functions dedicated to computing the cross-3PCF.

4.1 Cross-two-point correlation function monopole

Initially, we implemented new functions for computing cross correlations of the 2PCF monopole. The CBL already includes a function for computing a version of the cross-2PCF monopole, using the following expression (the one of Eq.(3.12)):

$$\hat{\xi}_{ij}(r) = \frac{(D_i - R)(D_j - R)}{RR} = \frac{D_iD_j(r) - D_iR(r) - D_jR(r) + RR(r)}{RR(r)}, \tag{4.1}$$

which employs two data catalogues, D_i and D_j, and a single random catalogue, R. We will denote this as the *single-random* method (SR). However, as discussed in Section

3.1.1, the SR method is inadequate for our objectives since it does not disentangle auto- and cross-signals. Instead, we require an estimator like the one defined in Eq.(3.11):

$$\hat{\xi}_{ij}(r) = \frac{(D_i - R_i)(D_j - R_j)}{R_i R_j} = \frac{D_i D_j(r) - D_i R_j(r) - D_j R_i(r) + R_i R_j(r)}{R_i R_j(r)}, \tag{4.2}$$

where two random catalogues (one for each data, using the same random-data ratio, D_R) are used. We refer to this as the *double-random* method (DR). The implementation of the DR cross-2PCF followed a template similar to the existing SR method. The main differences lie in pair counting and the final estimator formula.

Initially, we implemented a *count_allPairs_two_randoms* method to tally all pairs in Eq.(4.2) for each radial bin, utilizing two data catalogues, D_1 and D_2 (with N_{D_1} and N_{D_2} galaxies respectively), and two random catalogues, R_1 and R_2 (with N_{R_1} and N_{R_2} galaxies respectively). Subsequently, additional functions were developed for signal measurement. The non-normalized counts per redshift bin ($D_1{D_2}^*$, $D_1{R_2}^*$, $D_2{R_1}^*$, and $R_1{R_2}^*$) must be normalized in the following way:

$$\begin{aligned} D_1D_2(r) &= \frac{D_1{D_2}^*(r)}{N_{D_1}N_{D_2}}, \\ D_1R_2(r) &= \frac{D_1{R_2}^*(r)}{N_{D_1}N_{R_2}}, \\ D_2R_1(r) &= \frac{D_2{R_1}^*(r)}{N_{D_2}N_{R_1}}, \\ R_1R_2(r) &= \frac{R_1{R_2}^*(r)}{N_{R_1}N_{R_2}}. \end{aligned} \tag{4.3}$$

The apex * denotes the non-normalized number of couples. These normalized pair counts are then combined as in Eq.(4.2), and $\hat{\xi}_{12}$ is computed for each radial bin. Eventually, each $\hat{\xi}_{12}(r)$ is associated with a Poissonian error, $\Delta\xi_{12}(r)$, calculated as follows:

$$\begin{aligned} \Delta\xi_{12}^2(r) = &\left(\frac{N_{R_1}N_{R_2}}{N_{D_1}N_{D_2}}\frac{\sqrt{D_1{D_2}^*(r)}}{R_1{R_2}^*(r)}\right)^2 + \left(\frac{N_{R_1}}{N_{D_1}}\frac{\sqrt{D_1{R_2}^*(r)}}{R_1{R_2}^*(r)}\right)^2 + \left(\frac{N_{R_2}}{N_{D_2}}\frac{\sqrt{D_2{R_1}^*(r)}}{R_1{R_2}^*(r)}\right)^2 \\ &+ \left[\left(\frac{N_{R_1}N_{R_2}}{N_{D_1}N_{D_2}}D_1{D_2}^*(r) - \frac{N_{R_1}}{N_{D_1}}D_1{R_2}^*(r) - \frac{N_{R_2}}{N_{D_2}}D_2{R_1}^*(r)\right)\frac{1}{\sqrt{R_1{R_2}^*(r)}}\right]^2. \end{aligned} \tag{4.4}$$

The expressions for the Poissonian uncertainties of the estimators have been derived following the simple formula for the propagation of the uncertainties for a function $f(\boldsymbol{x})$

depending on i uncorrelated variables, x_i:

$$\Delta f(\boldsymbol{x})^2 = \sum_i \left(\frac{\partial f}{\partial x_i} \Delta x_i \right)^2 , \tag{4.5}$$

where Δx_i is the error associated to the variable x_i. In this case, the variables are the non-normalized number of couples and their associated errors are their Poissonian uncertainties.

4.1.1 Testing the cross-two-point correlation function monopole

In order to test the new functions, we conducted a direct comparison with existing, well-tested functions within the CBL. We compared the results of our new functions with the following two:

- The first function computes the auto-correlation of the 2PCF. It operates on a data catalogue D with N_D tracers and a random catalogue R with N_R tracers.
- The second function computes the cross-SR correlation of the 2PCF using the estimator defined in Eq.(4.1). It operates on two data catalogues D_1 and D_2 (with N_{D_1} and N_{D_2} tracers respectively) and one random catalogue R with N_R tracers.

For our testing, we utilized tracers only from one data and one random catalogue, thus setting $D_1 = D_2 = D$ and $R_1 = R_2 = R$. Under this assumption, both estimators for the cross-correlation (SR and DR, Eqs. (4.1) and (4.2) respectively) reduce to Eq.(3.3), the estimator for the auto-2PCF. Therefore, the signals from all three functions should be identical. Regarding the *non-normalized* pair counts, assuming $D_1 = D_2 = D$ and $R_1 = R_2 = R$, the rules are as follows:

- The auto-pairs counts, which originate from the same catalogue, should depend only on the size of the catalogues, in the same way across different estimators. Since in our case the catalogues are assumed to be identical, for instance the counts $RR^*(r)$ of the auto-estimator and the cross-SR estimator should be identical.
- The cross-pairs counts, involving tracers from different catalogues, should be doubled compared to auto-pairs from catalogues with equivalent size. This doubling occurs because cross-pairs are counted over all possible permutations of catalogues (in this case, 2). For example, the counts $R_1{R_2}^*(r)$ in the cross-DR estimator account for permutations R_1R_2 and R_2R_1, since R_1 and R_2 are considered different by the estimator, even though we know them to be identical in this case. Therefore, $R_1{R_2}^*(r)$ in the cross-DR estimator should be doubled compared to $RR^*(r)$ in the auto-estimator and cross-SR estimator.

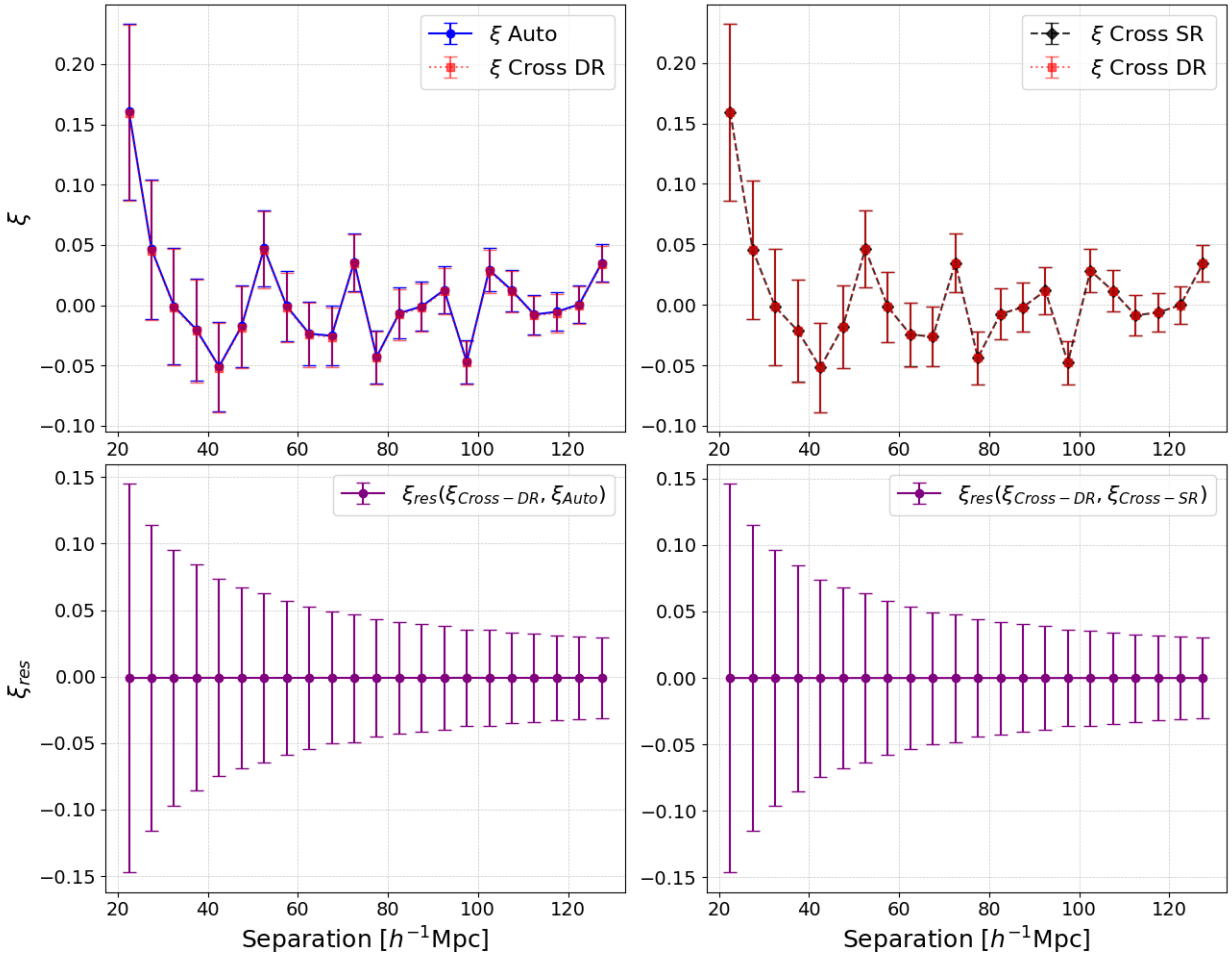


Figure 4.1: *Upper panels: results of the 2PCF monopole test. Measurements from the CBL test catalogue within a comoving separation range of* $20 - 130h^{-1}$Mpc. *The signals are calculated using the auto-estimator (solid blue line), Cross-SR-estimator (dashed black line), and Cross-DR-estimator (dotted red line). In order for the test to yield a positive outcome, the three functions should measure consistent quantities within the error bars (which represent Poissonian uncertainties). Lower panels: residuals between the auto-2PCF monopole and Cross-DR-2PCF monopole (left) and between the Cross-SR-2PCF monopole and Cross-DR-2PCF monopole signals (right). Error bars represent Poissonian uncertainties.*

- The cross-pairs counts should depend only on the size of the catalogues, in the same way across different estimators. That is since they are counted over the same number of permutations (2). For example, counts like $D_1R_2{}^*(r)$ (of cross-DR), $D_2R_1{}^*(r)$ (of cross-DR), $D_1R^*(r)$ (of cross-SR), $D_2R^*(r)$ (of cross-SR), and $DR^*(r)$ (of auto) should be identical.

The tests were conducted using a test catalogue from the CBL libraries. The latter

	DD	D_1D_2	DR	D_iR	D_iR_j	RR	R_1R_2
Auto-estimator	281	(none)	9643	(none)	(none)	95725	(none)
Cross-SR-estimator	(none)	562	(none)	9643	(none)	95725	(none)
Cross-DR-estimator	(none)	562	(none)	(none)	9643	(none)	191450

Table 4.1: *Number of pair counts for the first radial bin. The columns represent the kind of counted pairs, while each row corresponds to a different estimator.*

is a box-shaped catalogue containing 1110 galaxies. Its dimensions are 11.5 deg centered in 0 for both RA and Dec, and $\Delta z = 0.25$ centered in $z = 1$ for the redshift coordinate. Even though the number of tracers is too low to have significant statistics, it is still sufficient for testing purposes, as we only need to demonstrate that the signals of the different functions are equal, and ensures reasonable computation times. The random catalogue was constructed using a specific function of the CBL with a random-data ratio $D_R = 20$. Measurements were performed over a range from $r_{\min} = 20\,h^{-1}$Mpc to $r_{\max} = 120\,h^{-1}$Mpc, using 22 linear radial bins of width $\Delta r = 5\,h^{-1}$Mpc each. The testing results are summarized in Table 4.1 and Figure 4.1. Table 4.1 shows the number of each type of pairs for the first radial bin. It is easy to notice that the counts follow the rules described above. The two cross-pairs D_1D_2 in the cross-estimators are doubled compared to the auto-pairs DD, while DR pairs remain identical (all cross-pairs), and R_1R_2 in the cross-DR estimator is doubled compared to RR in the auto- and cross-SR estimators.

The upper panels of Figure 4.1 show that the signals from all three estimators are identical when $D_1 = D_2 = D$ and $R_1 = R_2 = R$.

This is corroborated by the bottom panels of Figure 4.1, which show the residuals $\xi_{res}(\xi_{Cross-DR}, \xi_{auto})$ in (left bottom panel) and $\xi_{res}(\xi_{Cross-DR}, \xi_{Cross-SR})$ (right bottom panel). The residual function, $\xi_{res}(\xi_i, \xi_j)$, is defined as follows:

$$\xi_{res}(\xi_i, \xi_j) \equiv \xi_i - \xi_j. \tag{4.6}$$

Both residuals are compatible with zero, as expected. This proves the well functioning of our new functions. The Poissonian errors of the residuals also behave as expected, being larger at smaller scales, where the number of counted pairs is lower.

4.2 Cross-two-point correlation function multipoles

Following the work on the monopole of the 2PCF, we implemented three classes and several functions to compute the first three even multipoles ($\ell = 0, 2, 4$, corresponding to the *monopole*, *quadrupole*, and *hexadecapole* respectively) of the Legendre expansion

of the estimator for the 2PCF (Eq.(3.6)):

$$\hat{\xi}^{int}_{ij,\ell}(r) = \frac{2\ell+1}{2}\int_{-1}^{+1}\left(1+\frac{D_1D_2(\mu,r)}{R_1R_2(\mu,r)}-\frac{D_1R_2(\mu,r)}{R_1R_2(\mu,r)}-\frac{D_2R_1(\mu,r)}{R_1R_2(\mu,r)}\right)P_\ell(\mu)\,\mathrm{d}\mu, \quad (4.7)$$

where D_1D_2, D_1R_2, D_2R_1, and R_1R_2 are the same defined in Eq.(4.3). However now they also depend on μ, the cosine of the angle between the comoving separation, r, and the line-of-sight. This double dependence requires 2D binning for the measurements in both comoving separation, r, and μ. The pairs were counted using the *count_allPairs_two_randoms* function discussed in the previous Section and were normalized in the same way. As with the previous computations, we associated a Poissonian error with the measurements of the multipoles of the 2PCF, following Eq.(4.4).

4.2.1 Testing the cross-two-point correlation function multipoles

The testing of these new functions have been conducted in the same manner as described in the previous Section. The CBL already includes validated functions for computing the monopole, quadrupole, and hexadecapole of the auto 2PCF multipoles. By assuming $D_1 = D_2 = D$ and $R_1 = R_2 = R$, the expressions in Eqs. (4.7) and (3.6) become identical. Consequently, we expect the signals from the auto- and auto-estimators to match if the measurements are performed with identical catalogues. As for the pair counts, we expect them to follow the behavior outlined in Paragraph 4.1.1.

The tests have been performed using the same data and random catalogues employed for testing the cross-2PCF monopole (Paragraph 4.1.1). Also in this case, measurements were conducted over a scale range from $r_{\min} = 20\,h^{-1}$Mpc to $r_{\max} = 120\,h^{-1}$Mpc, considering 22 linear radial bins each with a width of $\Delta r = 5\,h^{-1}$Mpc. The results of the tests are displayed in Figure 4.2, which upper panels show that the signals for the three multipoles from the two estimators are identical when $D_1 = D_2 = D$ and $R_1 = R_2 = R$. This is confirmed by the bottom panels of Figure 4.2, which show the residuals $\xi_{\ell,res}(\xi_{\ell,Cross}, \xi_{\ell,auto})$ for $\ell = 0, 2, 4$ (monopole, quadrupole and hexadecapole, respectively). The residuals are consistent with zero within the error bars, as expected. The Poissonian errors of the residuals behave as expected also in this case, being larger at smaller scales, where the number of counted pairs is lower.

4.3 Cross-three-point correlation function

The implementation of the cross-3PCF estimator presented the most challenging task among the three estimators. Our approach for this estimator diverged slightly from the method used for the cross-2PCF monopole and cross-2PCF multipoles. This time, the

	DD	D_1D_2	DR	D_iR_j	RR	R_1R_2
Auto-estimator	13	(none)	471	(none)	4872	(none)
Cross-estimator	(none)	26	(none)	471	(none)	9744

Table 4.2: *Number of pair counts for the first radial and angular bin. The columns represent the type of counted pairs, while each row corresponds to a different estimator.*

function not only computes the cross-correlation terms of the connected 3PCF but also includes the auto-correlation ones, allowing a comprehensive computation of all signals from the whole catalogue with a single function.

Initially, the *count_allTriplets* function was developed to count all types of triplets (DDD, DDR, DRR, and RRR) in a given triangular configuration (r_{12}, r_{13}, θ). Here we wanted to generalize the count to every possible catalogue combination described in Paragraph 3.3.1 ($i-i-i$, $i-i-j$, $i-j-j$, and $i-j-k$), resulting in the counting of 56 different kinds of triplets. Subsequently, another function normalizes the triplet counts for all possible combinations according to Eq.(3.23), that for three-point statistics becomes:

$$\begin{aligned} D_iD_iD_i &= \frac{D_iD_iD_i{}^*}{N_{D_i}(N_{D_i}-1)(N_{D_i}-2)}, \\ D_iD_iD_j &= \frac{D_iD_iD_j{}^*}{N_{D_i}(N_{D_i}-1)N_{D_j}}, \\ D_iD_jD_j &= \frac{D_iD_jD_j{}^*}{N_{D_i}N_{D_j}(N_{D_j}-1)}, \\ D_iD_jD_k &= \frac{D_iD_jD_k{}^*}{N_{D_i}N_{D_j}N_{D_k}}, \end{aligned} \tag{4.8}$$

and similarly for the other combinations of triplets. The function then aggregates these normalized counts to compute the signals of $\hat{\zeta}_i$, $\hat{\zeta}_{iij}$, $\hat{\zeta}_{ijj}$, and $\hat{\zeta}_{ijk}$. As we discussed in Paragraph 3.3.1, their expression is the following:

$$\hat{\zeta}_{iii} = \frac{D_iD_iD_i - 3D_iD_iR_i + 3D_iR_iR_i - R_iR_iR_i}{R_iR_iR_i}, \tag{4.9}$$

$$\hat{\zeta}_{ijj} = \frac{D_iD_iD_j - D_iD_iR_j - 2D_iR_iD_j + 2D_iR_iR_j + R_iR_iD_j - R_iR_iR_j}{R_iR_iR_j}, \tag{4.10}$$

$$\hat{\zeta}_{ijj} = \frac{D_iD_jD_j + D_iR_jR_j - 2D_iD_jR_j + 2R_iD_jR_j - R_iD_jD_j - R_iR_jR_j}{R_iR_jR_j}, \tag{4.11}$$

$$\hat{\zeta}_{ijk} = \frac{D_iD_jD_k - D_iR_jD_k - D_iD_jR_k - R_iD_jD_k + R_iR_jD_k + R_iD_jR_k + D_iR_jR_k - R_iR_jR_k}{R_iR_jR_k}. \tag{4.12}$$

Finally, a Poissonian error is associated with each measurement of the connected 3PCF to account for statistical uncertainties. The latter can be derived from Eq.(4.5) and can be written as follows for different types of 3PCF:

$$\Delta\zeta_i^2 = \left(\frac{N_{R_i}N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}N_{D_i}}\frac{\sqrt{D_iD_iD_i{}^*}}{R_iR_iR_i{}^*}\right)^2 + \left(3\frac{N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}}\frac{\sqrt{D_iD_iR_i{}^*}}{R_iR_iR_i{}^*}\right)^2 + \left(3\frac{N_{R_i}}{N_{D_i}}\frac{\sqrt{D_iR_iR_i{}^*}}{R_iR_iR_i{}^*}\right)^2 + \left[\left(\frac{N_{R_i}N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}N_{D_i}}D_iD_iD_i{}^* - 3\frac{N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}}D_iD_iR_i{}^* + 3\frac{N_{R_i}}{N_{D_i}}D_iR_iR_i{}^*\right)(R_iR_iR_i{}^*)^{-\frac{1}{2}}\right]^2, \tag{4.13}$$

$$\begin{aligned}\Delta\zeta_{iij}^2 = &\left(\frac{N_{R_i}N_{R_i}N_{R_j}}{N_{D_i}N_{D_i}N_{D_j}}\frac{\sqrt{D_iD_iD_j{}^*}}{R_iR_iR_j{}^*}\right)^2 + \left(\frac{N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}}\frac{\sqrt{D_iD_iR_j{}^*}}{R_iR_iR_j{}^*}\right)^2 \\ &+ \left(2\frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}\frac{\sqrt{D_iD_jR_i{}^*}}{R_iR_iR_j{}^*}\right)^2 + \left(2\frac{N_{R_i}}{N_{D_i}}\frac{\sqrt{D_iR_iR_j{}^*}}{R_iR_iR_j{}^*}\right)^2 + \left(\frac{N_{R_j}}{N_{D_j}}\frac{\sqrt{D_jR_iR_i{}^*}}{R_iR_iR_j{}^*}\right)^2 \\ &+ \left[\left(\frac{N_{R_i}N_{R_i}N_{R_j}}{N_{D_i}N_{D_i}N_{D_j}}D_iD_iD_j{}^* - \frac{N_{R_i}N_{R_i}}{N_{D_i}N_{D_i}}D_iD_iR_j{}^* - 2\frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}D_iD_jR_i{}^*\right.\right. \\ &\left.\left.+2\frac{N_{R_i}}{N_{D_i}}D_iR_iR_j{}^* - \frac{N_{R_j}}{N_{D_j}}D_jR_iR_i{}^*\right)(R_iR_iR_j{}^*)^{-\frac{1}{2}}\right]^2,\end{aligned} \tag{4.14}$$

$$\begin{aligned}\Delta\zeta_{ijj}^2 = &\left(\frac{N_{R_i}N_{R_j}N_{R_j}}{N_{D_i}N_{D_j}N_{D_j}}\frac{\sqrt{D_iD_jD_j{}^*}}{R_iR_jR_j{}^*}\right)^2 + \left(\frac{N_{R_j}N_{R_j}}{N_{D_j}N_{D_j}}\frac{\sqrt{D_jD_jR_i{}^*}}{R_iR_jR_j{}^*}\right)^2 \\ &+ \left(2\frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}\frac{\sqrt{D_iD_jR_j{}^*}}{R_iR_jR_j{}^*}\right)^2 + \left(2\frac{N_{R_j}}{N_{D_j}}\frac{\sqrt{D_jR_iR_j{}^*}}{R_iR_jR_j{}^*}\right)^2 + \left(\frac{N_{R_i}}{N_{D_i}}\frac{\sqrt{D_iR_jR_j{}^*}}{R_iR_jR_j{}^*}\right)^2 \\ &+ \left[\left(\frac{N_{R_i}N_{R_j}N_{R_j}}{N_{D_i}N_{D_j}N_{D_j}}D_iD_jD_j{}^* - \frac{N_{R_j}N_{R_j}}{N_{D_j}N_{D_j}}D_jD_jR_i{}^* - 2\frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}D_iD_jR_j{}^*\right.\right. \\ &\left.\left.+2\frac{N_{R_j}}{N_{D_j}}D_jR_iR_j{}^* - \frac{N_{R_i}}{N_{D_i}}D_iR_jR_j{}^*\right)(R_iR_jR_j{}^*)^{-\frac{1}{2}}\right]^2,\end{aligned} \tag{4.15}$$

$$\Delta\zeta^2_{ijk} = \left(\frac{N_{R_i}N_{R_j}N_{R_k}}{N_{D_i}N_{D_j}N_{D_k}}\frac{\sqrt{D_iD_jD_k{}^*}}{R_iR_jR_k{}^*}\right)^2 + \left(\frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}\frac{\sqrt{D_iD_jR_k{}^*}}{R_iR_jR_k{}^*}\right)^2 + \left(\frac{N_{R_i}N_{R_k}}{N_{D_i}N_{D_k}}\frac{\sqrt{D_iD_kR_j{}^*}}{R_iR_jR_k{}^*}\right)^2$$
$$+ \left(\frac{N_{R_j}N_{R_k}}{N_{D_j}N_{D_k}}\frac{\sqrt{D_jD_kR_i{}^*}}{R_iR_jR_k{}^*}\right)^2 + \left(\frac{N_{R_i}}{N_{D_i}}\frac{\sqrt{D_iR_jR_k{}^*}}{R_iR_jR_k{}^*}\right)^2 + \left(\frac{N_{R_j}}{N_{D_j}}\frac{\sqrt{D_jR_iR_k{}^*}}{R_iR_jR_k{}^*}\right)^2$$
$$+ \left(\frac{N_{R_k}}{N_{D_k}}\frac{\sqrt{D_kR_iR_j{}^*}}{R_iR_jR_k{}^*}\right)^2 + \left[\left(\frac{N_{R_i}N_{R_j}N_{R_k}}{N_{D_i}N_{D_j}N_{D_k}}D_iD_jD_k{}^* - \frac{N_{R_i}N_{R_j}}{N_{D_i}N_{D_j}}D_iD_jR_k{}^* - \frac{N_{R_i}N_{R_k}}{N_{D_i}N_{D_k}}D_iD_kR_j{}^*\right.\right.$$
$$\left.\left. - \frac{N_{R_j}N_{R_k}}{N_{D_j}N_{D_k}}D_jD_kR_i{}^* + \frac{N_{R_i}}{N_{D_i}}D_iR_jR_k{}^* + \frac{N_{R_j}}{N_{D_j}}D_jR_iR_k{}^*\frac{N_{R_k}}{N_{D_k}}D_kR_iR_j{}^*\right)(R_iR_jR_k{}^*)^{-\frac{1}{2}}\right]^2. \tag{4.16}$$

4.3.1 Testing the cross-three-point correlation function

To verify the functionality of the functions that count triplets and compute the cross-3PCF terms, we followed a similar methodology as in the previous Sections. Examining the expressions for the cross-signals of the connected 3PCF (Eqs. (4.10), (4.11), and (4.12)) it is straightforward to show that under the assumption $D_i = D_j = D_k = D$ and $R_i = R_j = R_k = R$, all these expressions reduce to Eq.(4.9), which is the signal of the auto-3PCF. This implies that all cross-signals computed with our new function, assuming $D_i = D_j = D_k = D$ and $R_i = R_j = R_k = R$, should match the classic connected 3PCF computed using the already validated functions of the CBL that employ a single data and a single random catalogue.

The logic for triplet counts follows the same principles as the pair counts discussed in Paragraph 4.1.1. Before proceeding with the discussion, it is essential to distinguish between triplet types and ζ types. The former are defined by the number of different catalogues from which the triplet components originate. For instance, auto-triplets involve one catalogue (e.g., $D_1D_1D_1$ or $R_2R_2R_2$), bi-cross-triplets involve two catalogues (e.g., $D_1R_2R_2$ or $R_1R_1R_3$), and tri-cross-triplets involve three catalogues (e.g., $D_1D_3R_2$ or $D_2R_2R_3$). On the other hand, ζ types refer to the contributions to the signals in Eqs. (4.9), (4.10), (4.11), and (4.12) (ζ_{iii}, ζ_{iij}, ζ_{ijj}, and ζ_{ijk}). Here, we distinguish sub-catalogues by their labels, (i, j, k), and do not differentiate between data and random components from the same catalogue. For example, a tri-cross-triplet $D_1D_2R_1$ contributes to the bi-cross-ζ ζ_{112}. Generally, auto-triplets only contribute to auto-ζ, bi-cross-triplets can contribute to both auto-ζ and bi-cross-ζ, while tri-cross-triplets can contribute to both bi-cross-ζ and tri-cross-ζ. The rules for the triplet counts are the following:

- If the triplets are of the same kind (auto- , bi-cross- or tri-cross-) their number depends only on the size of the catalogues, in the same way across different estimators. That is because if catalogues are identical, the number of the counts only

depends on the possible permutations of the triplet kind (1 for auto-triplets, 3 for bi-cross-triplets, 6 for tri-cross-triplets).

- The number ratio of different triplet types coming from catalogues with equivalent size corresponds to the ratio of the number of possible permutations. For example, the number of bi-cross-triplets is three times that of auto-triplets, and the number of tri-cross-triplets is twice that of bi-cross-triplets and six times that of auto-triplets.

The tests in this case were conducted using a catalogue from the Flagship2 simulation: a full octant of sky with a depth $\Delta z = 0.01$ centered at $z = 1$. The catalogue contains 337,818 *pure* galaxies (i.e. excluding redshift interlopers), and the random catalogue was constructed using a linear interpolation on the distribution in RA, Dec, and z of the data catalogue, preserving its exact geometry. The random-data ratio used was $D_R = 5$, sufficient for testing purposes and ensuring reasonable computational times. The choice of the catalogue for the testing was different with respect to the 2PCF cases since the 3PCF, being a three-point statistical tool, requires larger samples to provide significant results. For testing purposes, measurements were conducted using a fixed sides configuration ($r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 45\,h^{-1}$Mpc) with a radial bin size of $5\,h^{-1}$Mpc and 20 bins for the angular variable, θ. This configuration ensures rapid calculations and serves as an effective test for the newly implemented function.

The triplet counts for the first angular bin are displayed in Table 4.3. The counts follow the expectations, with tri-cross triplet numbers being twice the bi-cross ones, and bi-cross triplets being three times the auto-triplets from the auto-estimators. Exceptions are noted for DDR and DRR triplets, which have equal numbers in both auto- and bi-cross estimators since they remain bi-cross-triplets due to the different catalogues involved (i.e. D and R).

The signals of the 3PCF are shown in the left panels of Figures 4.3, 4.4, and 4.5. As expected, the signals are identical for all different estimators (i.e. ζ, ζ_{iii}, ζ_{iij}, ζ_{ijj}, and ζ_{ijk}, with $i, j, k = 1, \ldots, 3$). This must be true under the assumption $D_i = D_j = D_k = D$ and $R_i = R_j = R_k = R$. This equality is confirmed by the residual plots in the right panels of Figures 4.3, 4.4, and 4.5, which show the residual functions $\zeta_{\rm res}(\zeta_i, \zeta_j)$. All residuals are consistent with zero, confirming the accuracy of our implementation. The Poissonian errors of the residuals behave as expected, being larger at smaller scales of the third side of the triangle (i.e. for $\theta \sim 0, \pi$) for which the number of counted triplets is lower.

	DDD	DDR	DRR	RRR
Auto-ζ (OLD function)	1.17465e+05	1.45038e+06	6.56626e+06	1.09152e+07
Auto-ζ (NEW function)	1.17465e+05	1.45038e+06	6.56626e+06	1.09152e+07
Bi-cross-triplets from bi-cross-ζ	3.52395e+05	1.45038e+06	6.56626e+06	3.27456e+07
Tri-cross-triplets from bi-cross-ζ	(none)	2.90076e+06	1.31325e+07	(none)
Tri-cross-triplets from tri-cross-ζ	7.04790e+05	2.90076e+06	1.31325e+07	6.54912e+07

Table 4.3: *Triplet counts for the first angular bin. The columns represent the broad types of triplets: DDD, DDR, DRR, RRR. The rows differentiate between triplet types (auto-, bi-cross-, and tri-cross-) and ζ types (auto-, bi-cross-, and tri-cross-).*

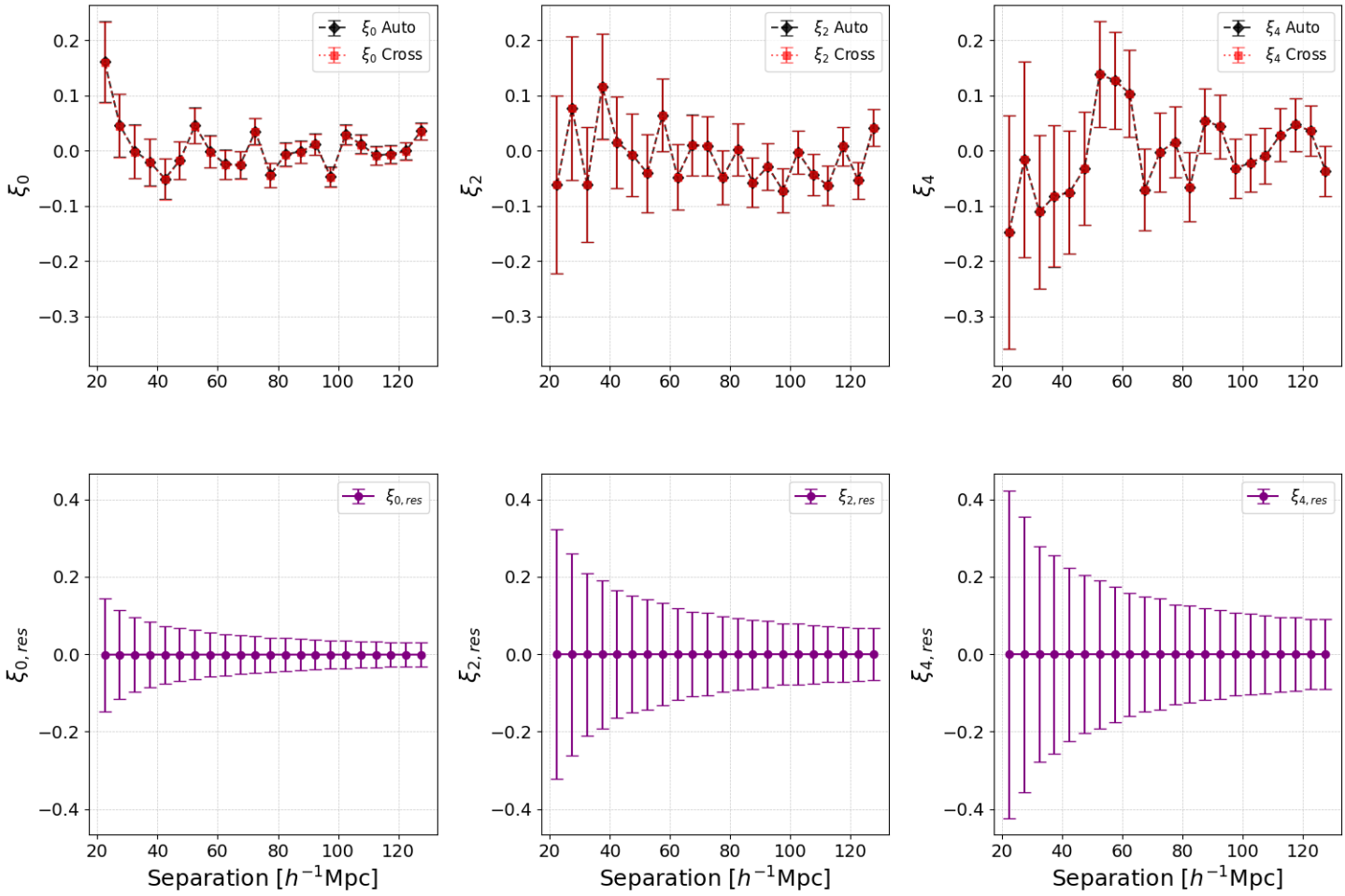
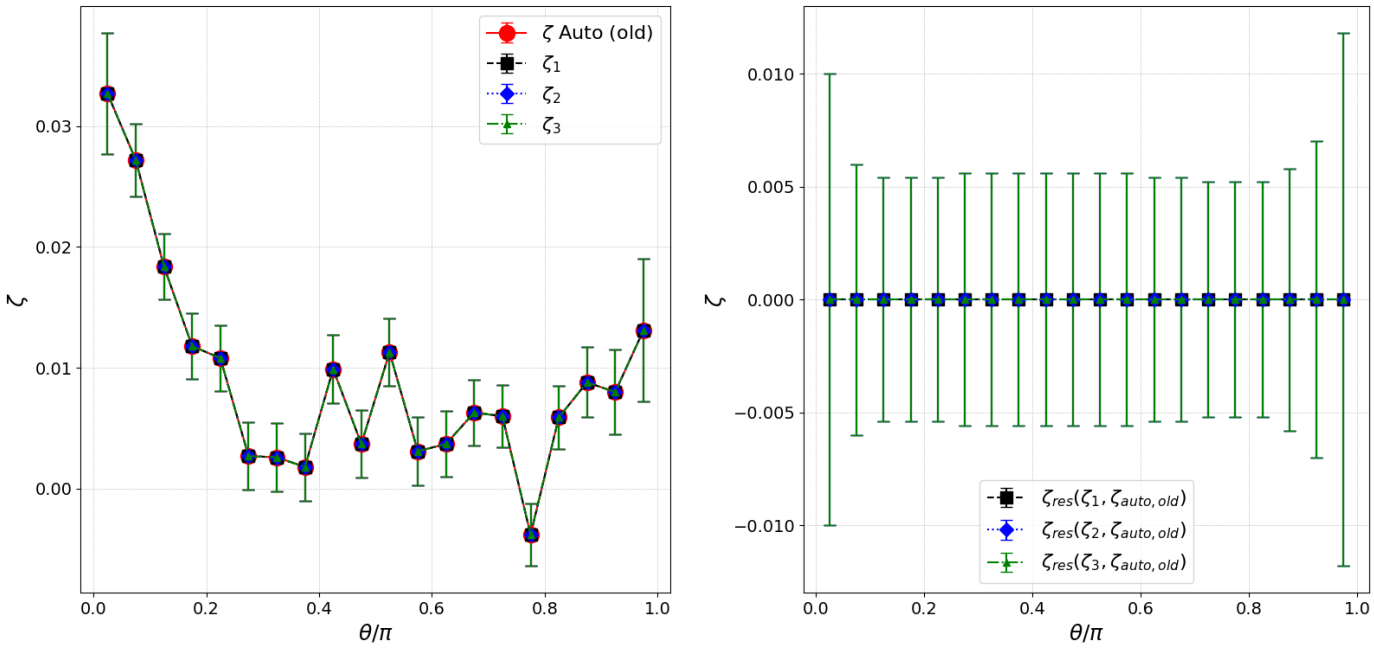
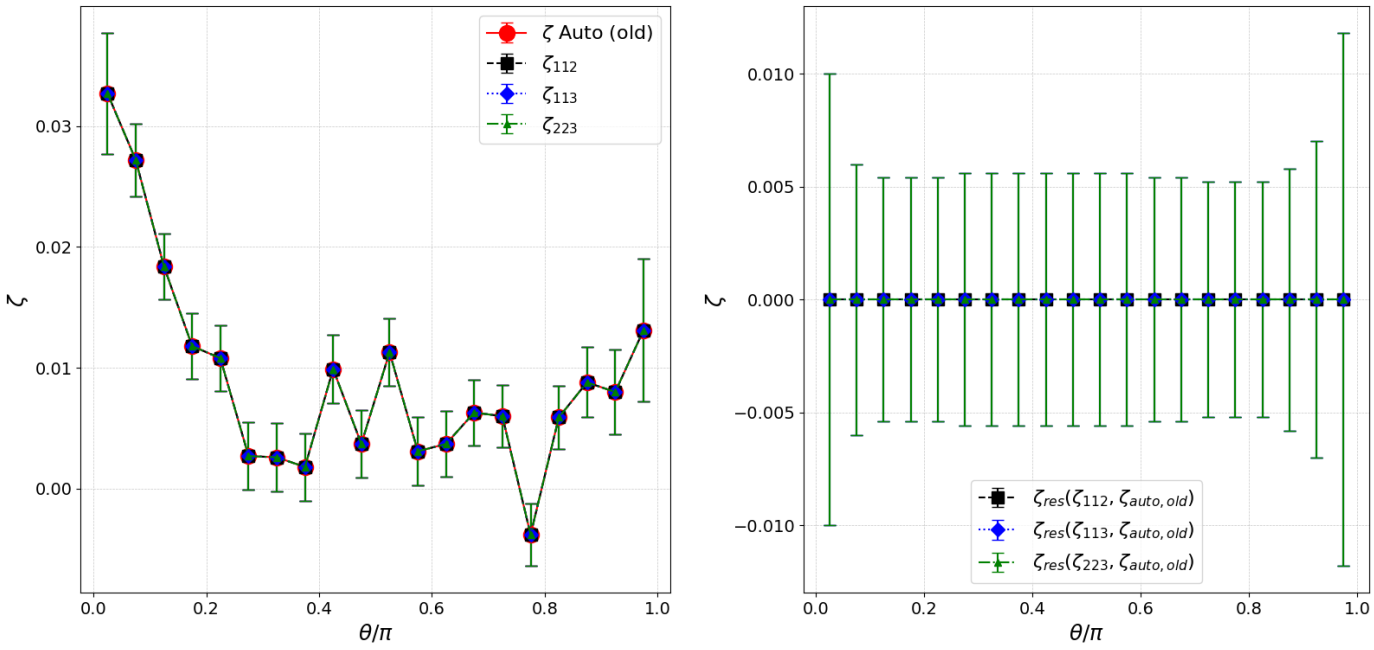


Figure 4.2: *Upper panels: 2PCF monopole (upper left), quadrupole (upper mid) and hexadecapole (upper right) measured from the CBL test catalogue within a comoving separation range of* $20 - 130h^{-1}$Mpc. *The signals are calculated using the auto-estimator (solid blue line) and the cross-estimator (dashed black line). Error bars represent Poissonian uncertainties. Lower panels: residuals of the 2PCF monopole (bottom left), quadrupole (bottom mid) and hexadecapole (bottom right) between the auto- and cross-2PCF signals. The latter are computed from the CBL test catalogue within a comoving separation range of* $20 - 130h^{-1}$*Mpc. Error bars represent Poissonian uncertainties.*



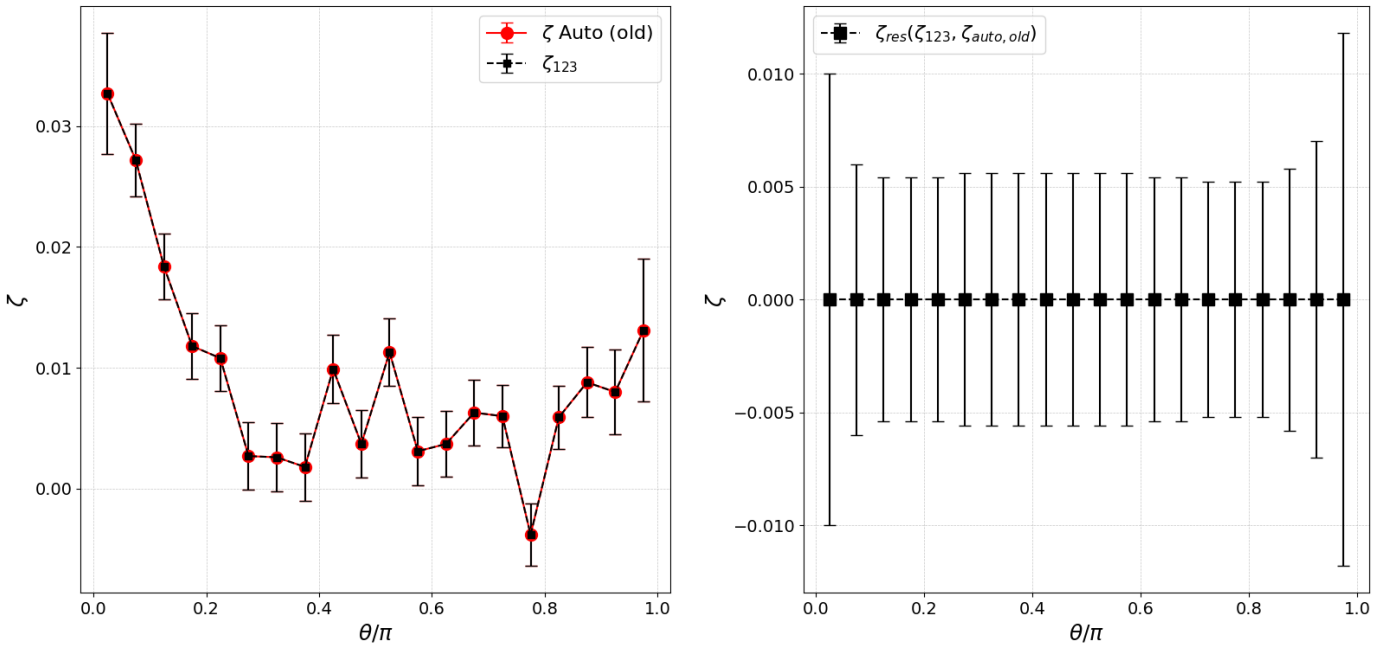
(a) (b)

Figure 4.3: *Left panel: the connected auto-3PCF, computed using the previously validated functions of the CBL (red continuous line), and the auto terms of the connected 3PCF, calculated using our newly developed functions, from the catalogues 1, 2 and 3 (black dashed line, blue dotted line and green dash-dotted line, respectively). The 3PCF was measured using the fixed-sides parametrization, with sides* $r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 45\,h^{-1}$Mpc). *Right panel: residuals between the three auto signals (from catalogues 1, 2 and 3) and the connected auto-3PCF. They are represented by a black dashed line, blue dotted line and green dash-dotted line, respectively. Poissonian error bars are applied to both plots.*



(a) (b)

Figure 4.4: *Left panel: the connected auto-3PCF, computed using the previously validated functions of the CBL (red continuous line), and the bi-cross terms of the connected 3PCF, calculated using our newly developed functions, between the catalogues 1-2, 1-3 and 2-3 (black dashed line, blue dotted line and green dash-dotted line, respectively). The 3PCF was measured using the fixed-sides parametrization, with sides* $r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 25\,h^{-1}$Mpc). *Right panel: residuals between the three bi-cross signals (from catalogues 1-2, 1-3 and 2-3) and the connected auto-3PCF. They are represented by a black dashed line, blue dotted line and green dash-dotted line, respectively. Poissonian error bars are applied to both plots.*



(a) (b)

Figure 4.5: *Left panel: the connected auto-3PCF, computed using the previously validated functions of the CBL (red continuous line), and the tri-cross term of the connected 3PCF, calculated using our newly developed functions (black dashed line). The 3PCF was measured using the fixed-sides parametrization, with sides* $r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 25\,h^{-1}$Mpc). *Right panel: residuals between the two signals presented in the left panel (black dashed line). Poissonian error bars are applied to both plots.*

Chapter 5

Measurements from the Flagship2 catalogue

In this Chapter, we will present the measurements of the 2PCF and 3PCF from the Flagship2 data set (details in Section 5.1). This analysis will be the first instance within this work to investigate the influence of line and noise interlopers on the signals of both the 2PCF and 3PCF. We will consider the auto- and cross-correlations of these interlopers and the pure galaxies. The computations were performed using the latest version of the CBL libraries [1].

5.1 Flagship2 catalogue

Euclid is a ESA mission, lauched in July 2022. In its wide survey, it will cover around $15000\,\mathrm{deg}^2$ of sky, measuring positions of galaxies up to redshift $z \sim 2$ and beyond. *Euclid* will exploit two instruments for the measurements: the Visible Instrument (VIS) and the Near-Infrared Spectrometer and Photometer (NISP). The former is an imager, which works at visible wavelengths, while the latter serves both as a photometer (NISP-p) and spectroscoper (NISP-s). The redshift of the *Euclid* wide survey will be estimated through spectroscopy based mainly on the $H\alpha$ line. Given the wavelength range covered by the NISP, the $H\alpha$ line is visible for the galaxies in the redshift range $0.84 < z < 1.98$.

The spectroscopic survey is expected to suffer from relatively high interloper contamination, i.e. systematic errors on the estimation of the galaxies' redshift due to line mismatch (for a detailed description see Section 2.4). This is due to the low SNR of the *Euclid* mission (the limiting SNR is 3.5 for a $F(H\alpha) = 2 \times 10^{-16}\,\mathrm{erg\,s^{-1}cm^2}$ line) and the reliance, in most cases, on a single emission line to estimate the spectroscopic redshift of the galaxies (i.e. $H\alpha$). This is expected to introduce both noise interlopers and line interlopers. In our study, we assume the line interlopers to be only OIII and SIII, as they are the strongest lines expected, after $H\alpha$, in the $0.84 < z < 1.98$ redshift range.

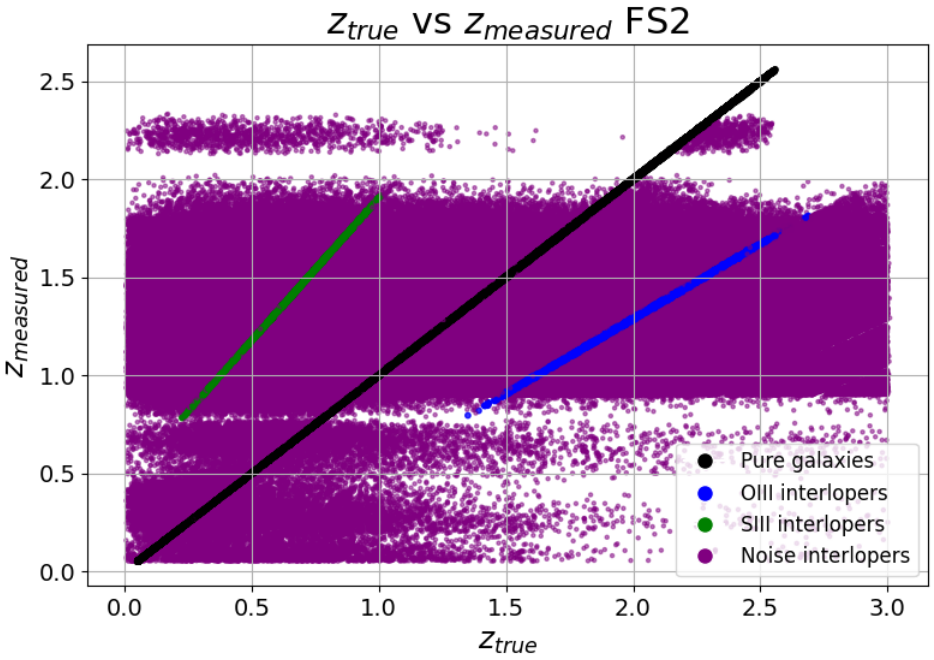


Figure 5.1: $z_{measured}$ *vs.* z_{true} *for the FS2 catalogue. The noise interlopers (purple dots) have a random distribution, the OIII and SIII interlopers (blue and green dots, respectively) lie one a line which angular coefficient R depends on the ratio between the rest-frame wavelength of their line and the H*α *line (Eq. 5.2). The pure galaxies (black dots) lie on the bisector of the I and III quadrant.*

Our goal in this Chapter is to study the 2PCF and 3PCF signal of the interlopers, and to quantify their effect on the 2PCF and 3PCF signals of the whole galaxy sample. For our purposes, we used a galaxy catalogue from Flagship2 (FS2), the *Euclid* official simulation, with manually inserted systematic redshift uncertainties. The catalogue was constructed within the Observational Systematic Work Package of the Science Working Group-Galaxy Clustering in *Euclid* using the following procedure. Two real galaxy catalogues were employed: COSMOS2020 and Emission Line COSMOS2020 (ELC2020). COSMOS2020 is a near-infrared selected galaxy catalogue containing 1.7 million galaxies with precise photometric redshifts (Weaver et al., 2022 [46]), serving as the reference for the *true* redshifts. ELC2020 is a catalogue in which emission lines are associated to COSMOS2020 galaxies, using an empirical but physically-motivated approach to model the galaxy emission-line fluxes in the COSMOS field (for details see Saito et al. 2020 [47]). Using the *Euclid*'s observational features, it simulates the survey and provides the *measured* redshifts, thus with possible interloper contamination, and predicts $H\alpha$ and OII galaxy number counts. In this way, we can associate both the *true* and the *measured* redshift (z_{true} and $z_{measured}$, respectively) to each galaxy of the catalogue. The relation

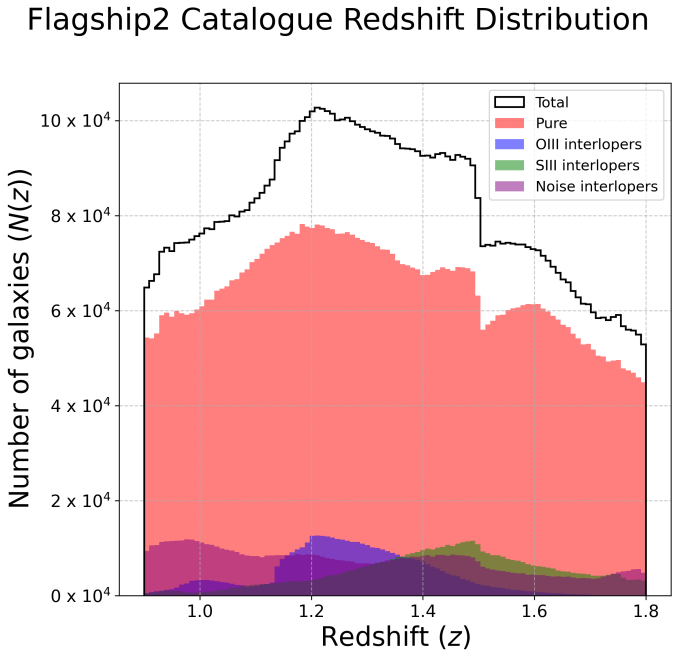


Figure 5.2: *Redshift distribution in the range* $0.9 < z < 1.8$ *for the total FS2 catalogue (solid black line) and the four sub-catalogues: pure (red), OIII (blue), SIII (green), and noise (purple).*

between z_{true} and $z_{measured}$ can be parametrized in this way:

$$z_{measured} = R \times (1 + z_{true}) - 1, \tag{5.1}$$

where R is defined as follows:

$$R \equiv \frac{\lambda_{e,true}}{\lambda_{e,measured}} = \frac{(1 + z_{measured})}{(1 + z_{true})}. \tag{5.2}$$

Here, λ_e is the rest-frame wavelength of the picked up line. The R coefficient behaves differently depending on the galaxy type:

- For *pure* galaxies, i.e. those whose redshift measurements are reliable since the line detected by the instrument is the correct one (Hα), $\lambda_{e,true} = \lambda_{e,measured}$, thus $R = 1$.
- For *line interlopers*, there is a fixed relation between the true rest-frame wavelength (OIII or SIII) and the measured rest-frame wavelength ($H\alpha$). Thus, $R =$ const, and $R \neq 1$.

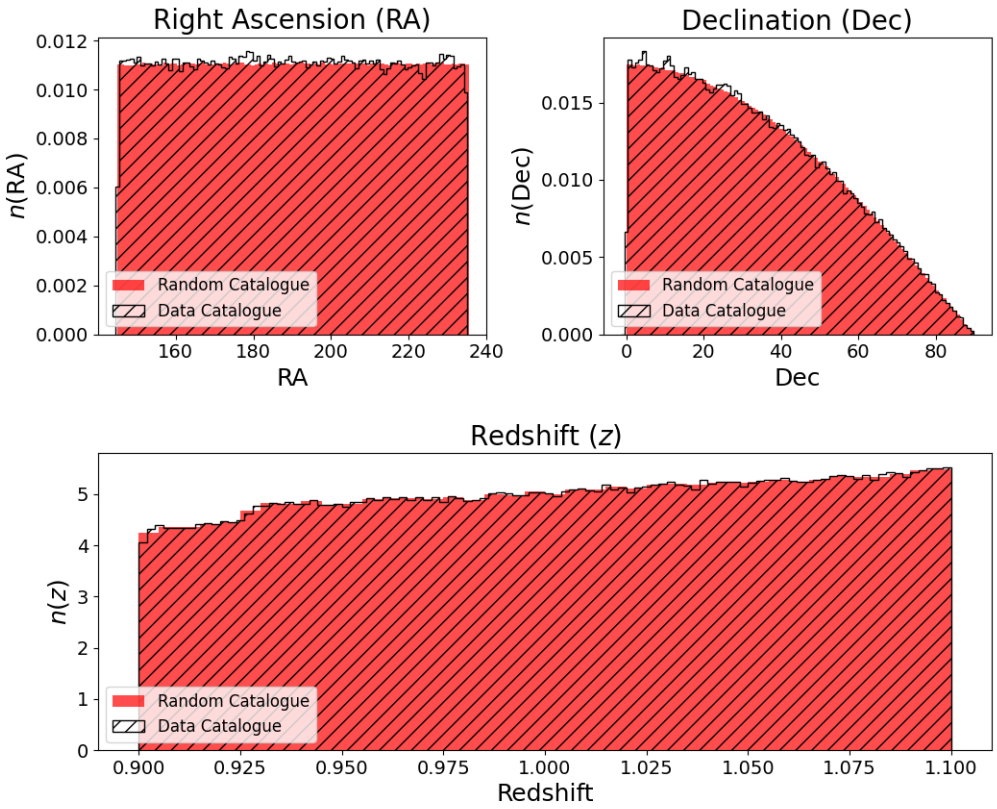


Figure 5.3: *Comparison of the Right Ascension, Declination, and redshift distributions ($n(RA)$, $n(Dec)$, and $n(z)$, respectively) of galaxies in the total FS2 catalogue (black line) and the corresponding random catalogue (red) within the first redshift range ($0.1 < z < 1.1$).*

- For *noise interlopers* there is *not* a fixed relation between the true rest-frame wavelength (a noise line) and the measured rest-frame wavelength ($H\alpha$). Thus, R has a random value.

As a matter of fact, from Eq.(5.1) it follows that $z_{measured} = f(z_{true})$ is the bisector of the z_{true} -$z_{measured}$ diagram for pure galaxies, a $y = R\,x + c$ line for the line interlopers and a random distribution in (x, y) for the noise interlopers. Knowing this, we can discriminate between the pure galaxies and the different kinds of interlopers, and flag them. Thus, by cross-correlating the two catalogues, it is possible to retrieve the redshift distribution, $n(z)$, of the interlopers and the systematic uncertainties can be inserted into the FS2 catalogue with the same distribution. Figure 5.1 shows the $z_{measured}$ vs.

z_{true} distribution for the FS2 catalogue galaxies. It is easy to identify the lines on which pure, OIII and SIII galaxies lie (as discussed before, their angular coefficient is R). The rest of the galaxies, which are randomly distributed in the plot, are noise interlopers. The FS2 galaxy catalogue replicates the typical redshift distribution of a *Euclid*-like catalogue and includes the expected percentage of interlopers in the *Euclid* survey, which is approximately 20% after the selection of the objects. As a matter of fact, if we select all the objects with magnitude $H_E < 24$ (the ones for which the *Euclid* survey will estimate the redshift), the catalogue would be dominated by interloper contamination, since there is a high percentage of galaxies which do not emit in $H\alpha$. In fact the *Euclid* requirements for the wide survey are 45% completeness and 80% purity, which will be reached with selections based on measured parameters such as redshift reliability and line flux. Thus, the vast majority of the galaxies will be *pure*, meaning their redshift measurements are reliable, that is the line detected by the instrument is the correct one (Hα).

The FS2 catalogue covers a full octant of the sky and includes most galaxies in the redshift range $0.9 < z < 1.8$. It contains the three observational coordinates of each galaxy in redshift space (right ascension, RA, declination, Dec, and redshift, z) and a label specifying their type (pure, OIII, SIII, or noise interloper), allowing the total catalogue to be easily divided into four sub-catalogues with a simple selection. The redshift distribution, $n(z)$, of the galaxies in the FS2 catalogue and sub-catalogues are displayed in Figure 5.2. The peculiar shape of the redshift distribution is related to the fact that the completeness of the survey strongly depends on the redshift. This is because of the previously discussed selection performed on the galaxies in order to achieve a 80% purity. For this reason, we observe redshift-dependent distributions for all types of galaxies except the noise interlopers, for which there is no fixed relation between the observed and the rest-frame wavelengths. In this work, we divided the FS2 catalogue into four redshift ranges to study them separately and gather information also on the evolution of the impact of interlopers with respect to cosmic time. The four ranges are: $0.9 < z < 1.1$, $1.1 < z < 1.3$, $1.3 < z < 1.5$, and $1.5 < z < 1.8$. The percentage composition of each redshift range is displayed in Table 5.1. As shown, the interlopers' percentages consistently add up to approximately 20%.

To compute the 2PCF and 3PCF, random samples are needed: we built them for the total FS2 catalogue and for each sub-catalogue in every redshift range, using a constant distribution in RA and sin(Dec), representing the geometry of a full octant of the sky. For the redshift distribution $n(z)$ of the random sample, we used a linear interpolator to match the distribution of the data catalogue. Figure 5.3 compares the RA, Dec, and z distributions of the total and random samples for the total FS2 catalogue in the first redshift range. The other 19 random catalogues used in this work (one for each redshift range of the total catalogue and the four sub-catalogues) have been constructed similarly.

In this case, the composition of the data and random catalogues (Eq. 3.8) can be

	Pure galaxies	OIII interlopers	SIII interlopers	Noise interlopers
$0.9 < z < 1.1$	81.8%	2.9%	1.5%	13.8%
$1.1 < z < 1.3$	78.2%	9.7%	3.9%	8.2%
$1.3 < z < 1.5$	74.5%	7.2%	9.9%	8.4%
$1.5 < z < 1.8$	83.4%	1.0%	8.6%	7.0%

Table 5.1: *Percentage composition of the FS2 catalogue in each redshift range. Rows show the redshift ranges, while columns distinguish between different sub-catalogues.*

expressed as follows:

$$\begin{aligned} D_{\rm tot} &= D_p + D_o + D_s + D_n, \\ R_{\rm tot} &= R_p + R_o + R_s + R_n, \end{aligned} \tag{5.3}$$

where the subscripts p, o, s, and n indicate the pure, OIII interlopers, SIII interlopers, and noise interloper sub-catalogues, respectively. Considering Eq.(5.3), we can derive the expression for the composition of the total 2PCF, $\xi_{\rm tot}$, and 3PCF, $\zeta_{\rm tot}$ (Eqs. 3.15 and 3.33, respectively) in the case of the FS2 catalogue. From Eq.(3.15), we obtain the following expression for $\xi_{\rm tot}$:

$$\begin{aligned} \xi_{\rm tot} =& f_p^2 \frac{R_p R_p}{RR}\xi_p + f_o^2 \frac{R_o R_o}{RR}\xi_o + f_s^2 \frac{R_s R_s}{RR}\xi_s + f_n^2 \frac{R_n R_n}{RR}\xi_n \\ &+ 2 f_p f_o \frac{R_p R_o}{RR}\xi_{po} + 2 f_p f_s \frac{R_p R_s}{RR}\xi_{ps} + 2 f_p f_n \frac{R_p R_n}{RR}\xi_{pn} \\ &+ 2 f_o f_s \frac{R_o R_s}{RR}\xi_{os} + 2 f_o f_n \frac{R_o R_n}{RR}\xi_{on} + 2 f_s f_n \frac{R_s R_n}{RR}\xi_{sn}, \end{aligned} \tag{5.4}$$

Similarly, from Eq.(3.33), we obtain the following expression for $\zeta_{\rm tot}$:

$$\begin{aligned}\zeta_{\rm tot} =& f_p^3\frac{R_pR_pR_p}{RRR}\zeta_{ppp} + f_o^3\frac{R_oR_oR_o}{RRR}\zeta_{ooo} + f_s^3\frac{R_sR_sR_s}{RRR}\zeta_{sss} + f_n^3\frac{R_nR_nR_n}{RRR}\zeta_{nnn}\\ &+3\left(f_p^2f_o\frac{R_pR_pR_o}{RRR}\zeta_{ppo} + f_pf_o^2\frac{R_pR_oR_o}{RRR}\zeta_{poo}\right)\\ &+3\left(f_p^2f_s\frac{R_pR_pR_s}{RRR}\zeta_{pps} + f_pf_s^2\frac{R_pR_sR_s}{RRR}\zeta_{pss}\right)\\ &+3\left(f_p^2f_n\frac{R_pR_pR_n}{RRR}\zeta_{ppn} + f_pf_n^2\frac{R_pR_nR_n}{RRR}\zeta_{pnn}\right)\\ &+3\left(f_o^2f_s\frac{R_oR_oR_s}{RRR}\zeta_{oos} + f_of_s^2\frac{R_oR_sR_s}{RRR}\zeta_{oss}\right)\\ &+3\left(f_o^2f_n\frac{R_oR_oR_n}{RRR}\zeta_{oon} + f_of_n^2\frac{R_oR_nR_n}{RRR}\zeta_{onn}\right)\\ &+3\left(f_s^2f_n\frac{R_sR_sR_n}{RRR}\zeta_{ssn} + f_sf_n^2\frac{R_sR_nR_n}{RRR}\zeta_{snn}\right)\\ &+6f_pf_of_s\frac{R_pR_oR_s}{RRR}\zeta_{pos} + 6f_pf_of_n\frac{R_pR_oR_n}{RRR}\zeta_{pon}\\ &+6f_pf_sf_n\frac{R_pR_sR_n}{RRR}\zeta_{psn} + 6f_of_sf_n\frac{R_oR_sR_n}{RRR}\zeta_{osn}.\end{aligned} \tag{5.5}$$

5.2 Two-point correlation function measurements

The composition of the FS2 catalogue allows us to investigate the impact of interlopers on the results, through the calculation of all terms in Eqs. (5.4) and (5.5). We begin by measuring the auto-correlations within the sub-catalogues to facilitate the estimation of the impacts of the cross-terms through residual analysis, i.e. subtracting the auto-correlation terms from the total signal.

We started from the multipoles of the auto-2PCF (ξ_p, ξ_o, ξ_s, and ξ_n) and of the total contaminated catalogue, $\xi_{\rm tot}$, across the four redshift ranges: $0.9 < z < 1.1$, $1.1 < z < 1.3$, $1.3 < z < 1.5$, and $1.5 < z < 1.8$. The measurements were performed using the direct LS estimator for the multipoles of the 2PCF (Eq. 3.7), within the comoving separation range $20 - 180\,h^{-1}$Mpc, using a linear binning of $\Delta r = 5\,h^{-1}$Mpc. The used random-data ratio was $D_R = 30$. The comparison of the monopole, and quadrupole signals of the 2PCFs is displayed in Figures 5.4 and 5.5.

These initial results align with our expectations. The signal from SIII interlopers is higher than the pure signal, as mistaking SIII for $H\alpha$ results in an overestimation of the tracers' redshift. This indicates that SIII interlopers actually belong to a lower redshift bin than pure galaxies, thereby increasing the 2PCF signal (since it resembles the one of a more evolved universe). Conversely, the auto-signal from OIII interlopers is lower

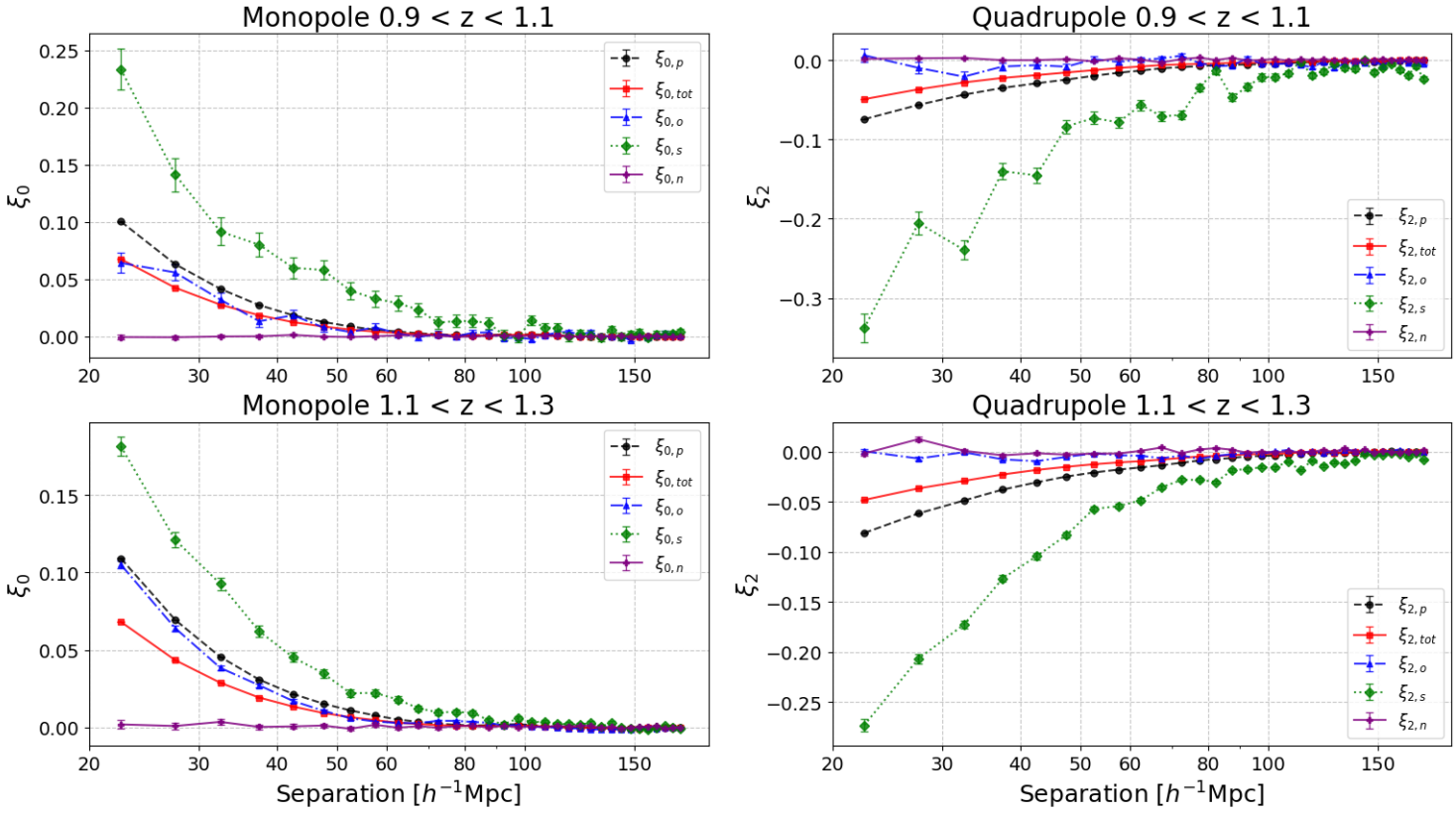


Figure 5.4: *Comparison between the total 2PCF signals (solid red line) and the auto-2PCF of pure galaxies (black dashed line), OIII interlopers (blue dash-dotted line), SIII interlopers (green dotted line), and noise interlopers (purple solid line) within a comoving separation range of* $20 - 180h^{-1}$ *Mpc. The upper panels show the signal in the first redshift bin* ($0.9 < z < 1.1$) *and the lower panels show the signal in the second redshift bin* ($1.1 < z < 1.3$). *The left panels show the monopole of the 2PCF,* ξ_0, *while the right panels show the quadrupole of the 2PCF,* ξ_2. *Error bars represent Poissonian uncertainties. The x-axis is in logarithmic scale.*

than the pure signal, resembling that of a higher redshift bin. Noise interlopers show no significant signal, which is expected because there is no fixed relation between the wavelength of $H\alpha$ and noise lines picked up by the instrument. This contrasts with SIII and OIII interlopers, as the noise population resembles a random one and thus has no detectable 2PCF signal.

The most notable finding is the overall attenuation of the total signal compared to the

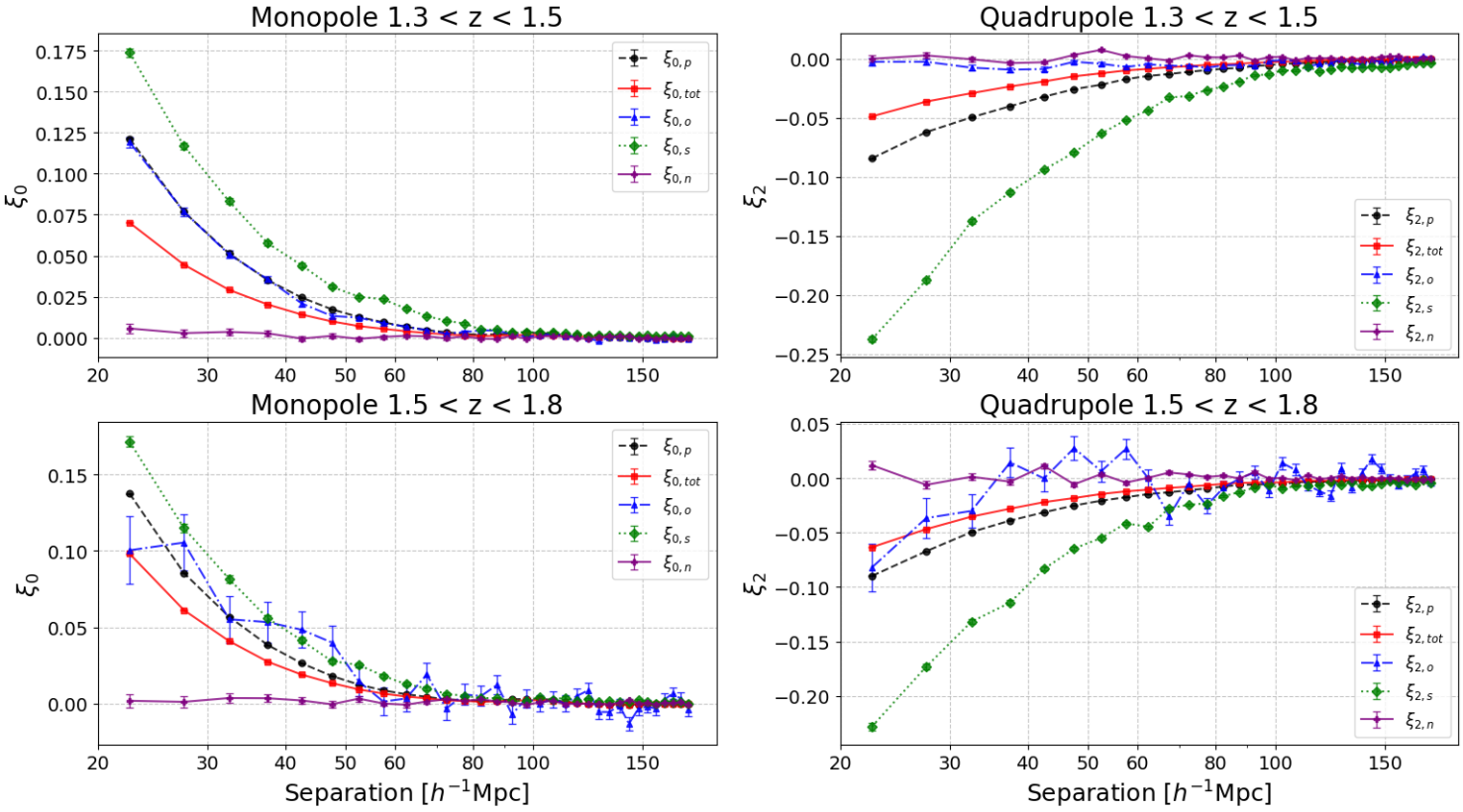


Figure 5.5: *Comparison between the total 2PCF signals (solid red line) and the auto-2PCF of pure galaxies (black dashed line), OIII interlopers (blue dash-dotted line), SIII interlopers (green dotted line), and noise interlopers (purple solid line) within a comoving separation range of* $20 - 180h^{-1}$ *Mpc. The upper panels show the signal in the third redshift bin* ($1.3 < z < 1.5$) *and the lower panels show the signal in the fourth redshift bin* ($1.5 < z < 1.8$). *The left panels show the monopole of the 2PCF,* ξ_0, *while the right panels show the quadrupole of the 2PCF,* ξ_2. *Error bars represent Poissonian uncertainties. The x-axis is in logarithmic scale.*

pure signal due to contamination by interlopers. Table 5.2 shows the average damping of the monopole signal due to interloper contamination at small scales ($20 - 120\,h^{-1}$Mpc) and large scales ($150 - 180\,h^{-1}$Mpc) across the four redshift ranges. The values exceed 25%, touching peaks above 75%. Moreover, we notice that the attenuation is lower in the fourth redshift range, suggesting that the impact of the interlopers is slightly weakened above redshift $z = 1.5$. Summing up, the values suggest that the 2PCF signal is strongly

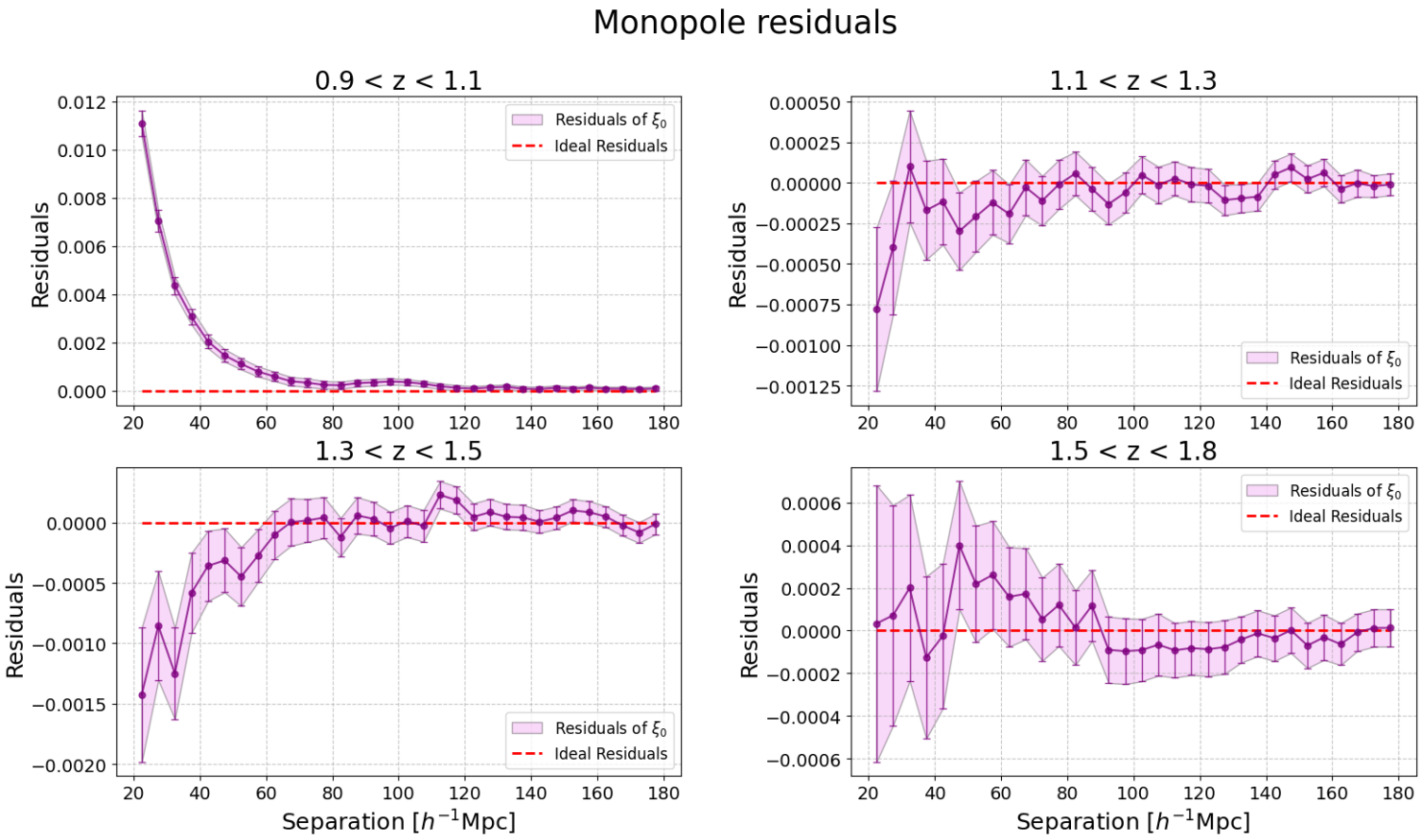


Figure 5.6: *Residuals of the total 2PCF monopole signal of the FS2 catalogue and the monopole auto-signals of its sub-catalogues. The residuals are computed using Eq.*(5.6)*. The ideal residuals (zero) are indicated with a red dashed line. Each panel represents a different redshift range. Error bars represent Poissonian uncertainties.*

affected by interloper contamination in all redshift ranges.

We then wanted to estimate the impact of the cross-signals on the total 2PCF. As previously mentioned, we did that by performing a residual analysis. Figure 5.6 presents the residuals of the total 2PCF monopole and the auto-signal monopole of the sub-catalogues, expressed as:

$$\xi_{0,res} = \xi_{0,\mathrm{tot}} - f_p^2 \frac{R_p R_p}{RR}\xi_{0,p} - f_o^2 \frac{R_o R_o}{RR}\xi_{0,o} - f_s^2 \frac{R_s R_s}{RR}\xi_{0,s} - f_n^2 \frac{R_n R_n}{RR}\xi_{0,n}. \tag{5.6}$$

As shown in the plots, the residuals are not always consistent with zero, indicating that cross-correlations cannot be ignored when recovering the complete 2PCF signal,

	Average $\frac{\xi_{0,p}-\xi_{0,tot}}{\xi_{0,p}}$ (small scales)	Average $\frac{\xi_{0,p}-\xi_{0,tot}}{\xi_{0,p}}$ (large scales)
$0.9 < z < 1.1$	26.2%	78.3%
$1.1 < z < 1.3$	37.6%	37.4%
$1.3 < z < 1.5$	39.7%	57.3%
$1.5 < z < 1.8$	27.8%	25.5%

Table 5.2: *Average percentual damping of the 2PCF monopole coming from the total FS2 catalogue, $\xi_{0,tot}$, with respect to the one coming from the pure sub-catalogue, $\xi_{0,p}$. The values are reported for small scales ($20 - 120\,h^{-1}$Mpc) in the first column and for large scales ($150 - 180\,h^{-1}$Mpc) in the second column. Each row refers to a different redshift range.*

particularly at small scales. If we quantify the impact of the residuals on the total 2PCF signal for all redshift ranges in the comoving separation range $20-120\,h^{-1}$Mpc, we notice that the importance of cross-correlations fades at higher redshift. As a metter of fact, the residuals are around 25% of the total signal in the first redshift range, and they lower to 10% and 5% in the second and third range, respectively. For $1.5 < z < 1.8$ the impact of cross-correlations is almost null: this is also visible in the bottom right panel of Fig. 5.6, where the residuals are consistent with zero. We expect then that higher-redshift measurements will be less affected by interloper cross-correlations.

The most peculiar behavior is observed in the first redshift range (upper left panel of Fig. 5.6), which resembles a genuine 2PCF signal. This phenomenon is attributed to significant contamination by noise interlopers in the first redshift bin. As detailed in Table 5.1, the fractions of OIII and SIII interlopers in the first redshift range are low (2.9% and 1.5%, respectively), resulting in very small pre-factors in Eq.(5.6) (of the order of 10^{-4}), thereby making their contributions to the residuals negligible. The noise interlopers, as discussed earlier, do not exhibit a signal, making their contribution to Eq.(5.6) also negligible ($\xi_{0,n} \sim 0$). Consequently, the residuals in the first redshift bin are essentially the difference between the total monopole signal and the rescaled auto-pure monopole signal, as the only significant terms in the right-hand-side of Eq.(5.6) are the first two. This results in the pattern observed in the upper left panel of Fig. 5.6.

In summary, to recover the total 2PCF signal, the auto-correlations of the sub-catalogues are insufficient. It is essential to compute their cross-correlations. This task will be addressed in the following Section, using the functions discussed in Chapter 4.

5.2.1 Cross-correlations of the two-point correlation function

After analyzing the auto-2PCF, we proceeded with the cross-correlations, quantifying their role in recovering the total 2PCF signal from the FS2 catalogue. This is particularly crucial since, as discussed in the last Paragraph, we found that the cross-correlation

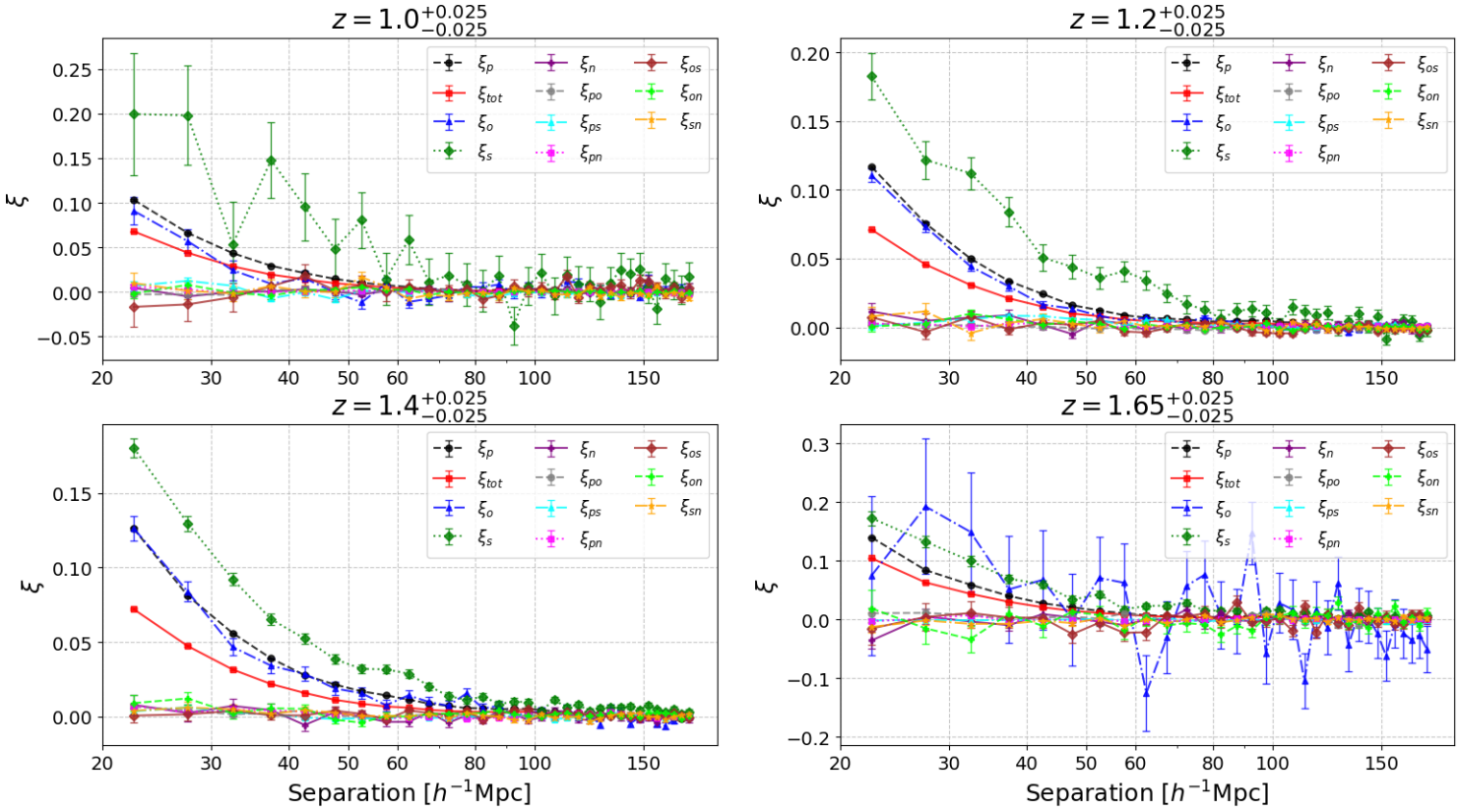


Figure 5.7: *Comparison between the total 2PCF monopole signals (solid red line), the auto-2PCF monopole of pure galaxies (black dashed line), OIII interlopers (blue dash-dotted line), SIII interlopers (green dotted line), and noise interlopers (purple solid line), and the cross-2PCF monopole signals within a comoving separation range of* $20-180h^{-1}$ *Mpc. Each panel represents a different redshift range. Error bars represent Poissonian uncertainties. The x-axis is in logarithmic scale.*

can contribute up to 25% of the total signal. Considering computational constraints, cross-2PCF measurements were conducted in redshift slices of width $\delta z = 0.5$, centered at $z = 1.0, 1.2, 1.4, 1.65$, within each of the four specified redshift ranges. The same random and data catalogues used previously were employed, with a specific redshift coordinate selection applied. For the cross-2PCF measurement, we utilized the cross-SR estimator for the monopole, as detailed in Section 4.1, ensuring consistent binning with the auto-correlations. The measurements were taken in the comoving separation range $20-180\,h^{-1}$Mpc, using a binning interval of $\Delta r = 5\,h^{-1}$Mpc. We adopted a random-data

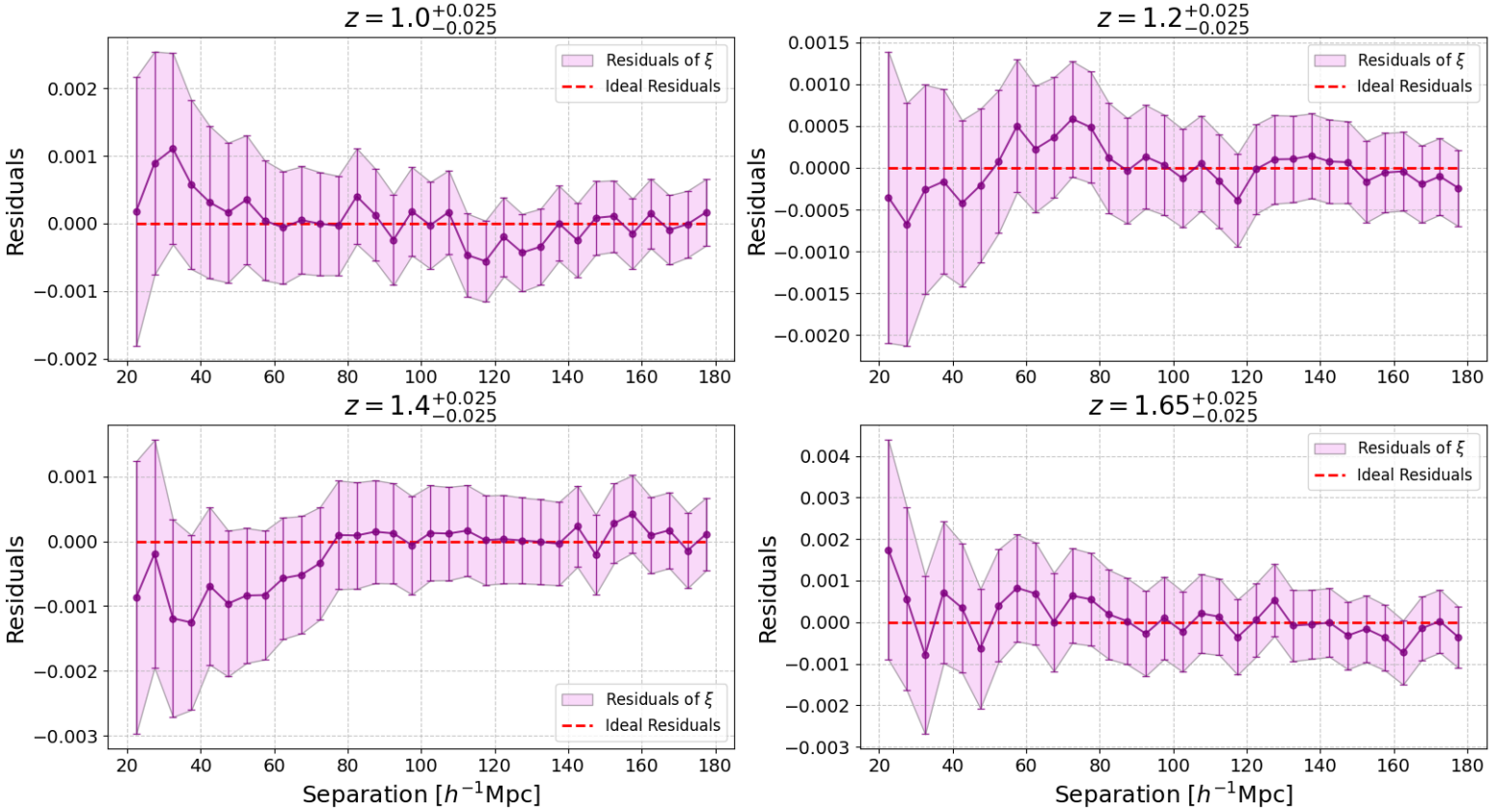


Figure 5.8: *Complete residuals (Eq. 5.7) of the total 2PCF monopole signal of the FS2 catalogue and the monopole auto-signals of its sub-catalogues. The ideal residuals (zero) are indicated with a red dashed line. Each panel represents a different redshift range. Error bars represent Poissonian uncertainties.*

ratio of $D_R = 5$. The resulting auto- and cross-2PCF signals for each redshift slice are displayed in Figure 5.7. The auto-signals exhibit analogous behavior to that described in the last Paragraph across all redshift ranges, albeit with larger Poissonian errors due to smaller catalogue sizes. The larger scatter in the SIII and OIII signal in the first and fourth redshift range, respectively, is due to the shortage of such interlopers in the corresponding ranges (see Table 5.1).

Conversely, the cross-signals do not reveal distinctive features. Their magnitude is approximately one order of magnitude lower than that of the auto-correlations, comparable instead to the auto-noise signal. Notably, these signals indicate a preference for small comoving separations, characterized by positive correlations at shorter scales.

This aligns with expectations, as there is no compelling reason for galaxies from different sub-catalogues to exhibit strong correlations. However, fluctuations in the cross-signals impact the total 2PCF measurements of the FS2 catalogue, in particular those in which the pure sample is involved (e.g. ξ_{pn}). This is because the value of the cross-correlation pre-factors of Eq.(5.4) can be much higher than the ones of the auto-correlations. For example, in the first redshift range $f_o^2 \sim 8 \times 10^{-4}$ and $f_p f_o \sim 2.4 \times 10^{-2}$.

Incorporating also the cross-signals into the residual computations yields expected outcomes. The residuals of the total 2PCF, including both auto- and cross-correlations, can be expressed as:

$$\begin{aligned}\xi_{0,res} = \xi_{0,\mathrm{tot}} &- f_p^2 \frac{R_p R_p}{RR}\xi_{0,p} - f_o^2 \frac{R_o R_o}{RR}\xi_{0,o} - f_s^2 \frac{R_s R_s}{RR}\xi_{0,s} - f_n^2 \frac{R_n R_n}{RR}\xi_{0,n} \\ &- 2 f_p f_o \frac{R_p R_o}{RR}\xi_{0,po} - 2 f_p f_s \frac{R_p R_s}{RR}\xi_{0,ps} - 2 f_p f_n \frac{R_p R_n}{RR}\xi_{0,pn} \\ &- 2 f_o f_s \frac{R_o R_s}{RR}\xi_{0,os} - 2 f_o f_n \frac{R_o R_n}{RR}\xi_{0,on} - 2 f_s f_n \frac{R_s R_n}{RR}\xi_{0,sn}.\end{aligned} \tag{5.7}$$

As shown in Figure 5.8, the residuals consistently approach zero across all four redshift slices. This means that the cross-signals fully cover the difference between the total signal and the auto-signals of the sub-catalogues, proving the validity of Eq.(3.15).

5.3 Three-point correlation function measurements

Following our analysis of the 2PCF, we proceeded to examine the 3PCF signals, beginning with the auto-signals as defined in Eq.(5.5), across the four redshift ranges: $0.9 < z < 1.1$, $1.1 < z < 1.3$, $1.3 < z < 1.5$, and $1.5 < z < 1.8$. The multipoles of the auto-3PCFs were measured up to $\ell = 11$ for all the fixed side configurations with r_{12} and r_{13} spacing in the range of $15 - 165\,h^{-1}$Mpc, using a binning interval of $\Delta r = 10\,h^{-1}$Mpc. The used random-data ratio was $D_R = 30$. This measurement was carried out using a CBL function that utilizes the SHD method as proposed by Slepian and Eisenstein (2015) [40]. Subsequently, the multipoles were combined to reconstruct the signal of the connected 3PCF in the two fixed sides configuration for all side combinations within the $15 - 165\,h^{-1}$Mpc range, using 20 bins for the angular variable. It is important to note that the CBL does not currently implement a function for computing the errors of the 3PCF multipoles. Therefore, the results presented here are shown without uncertainties. This issue will be addressed in the next Chapter, where we will estimate the covariance matrix using mock catalogues.

Figure 5.9 presents a comparison between the all-scales signal derived from the total F2 catalogue and that from the pure sub-catalogue across the four redshift ranges. In the plot, the triangle index refers to the number of the triangular configuration, which is characterized by $20\,h^{-1}\mathrm{Mpc} < r_{12} < r_{13} < r_{23} < 120\,h^{-1}\mathrm{Mpc}$, with a radial bin of

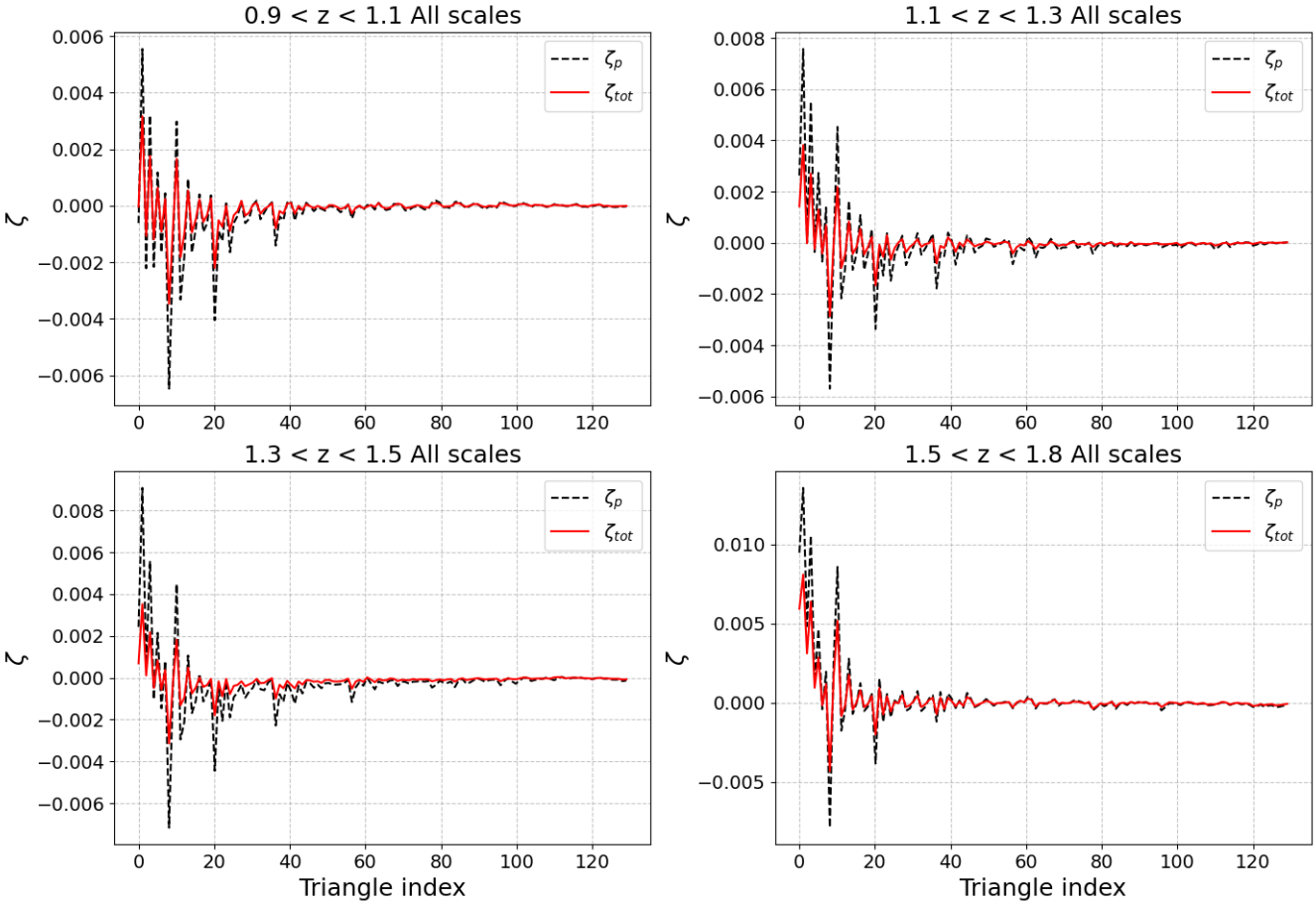


Figure 5.9: *Comparison between the total (black dashed line) and pure (red solid line) signal of the all-scales connected 3PCF. The triangle index refers to the number of the triangular configuration, which is characterized by* $20\,h^{-1}Mpc < r_{12} < r_{13} < r_{23} < 120\,h^{-1}Mpc$*, with a radial bin of* $\Delta r = 10\,h^{-1}Mpc$*. The configurations are organized in increasing order of their side lengths. Each panel represents a different redhsift bin:* $0.9 < z < 1.1$ *(upper left),* $1.1 < z < 1.3$ *(upper right),* $1.3 < z < 1.5$ *(lower left) and* $1.5 < z < 1.8$ *(lower right).*

$\Delta r = 10\,h^{-1}$Mpc. The configurations are organized in increasing order of their side lengths. Furthermore, all triangles with $\eta < 3$ are excluded, where η is defined as:

$$\eta = \frac{r_{13} - r_{12}}{\Delta r}. \tag{5.8}$$

This selection was performed to exclude equilateral and isosceles configurations, which require a higher number of multipoles to be accurately measured using the SHD method. We observed that the total all-scales 3PCF signal is weaker compared to the pure one

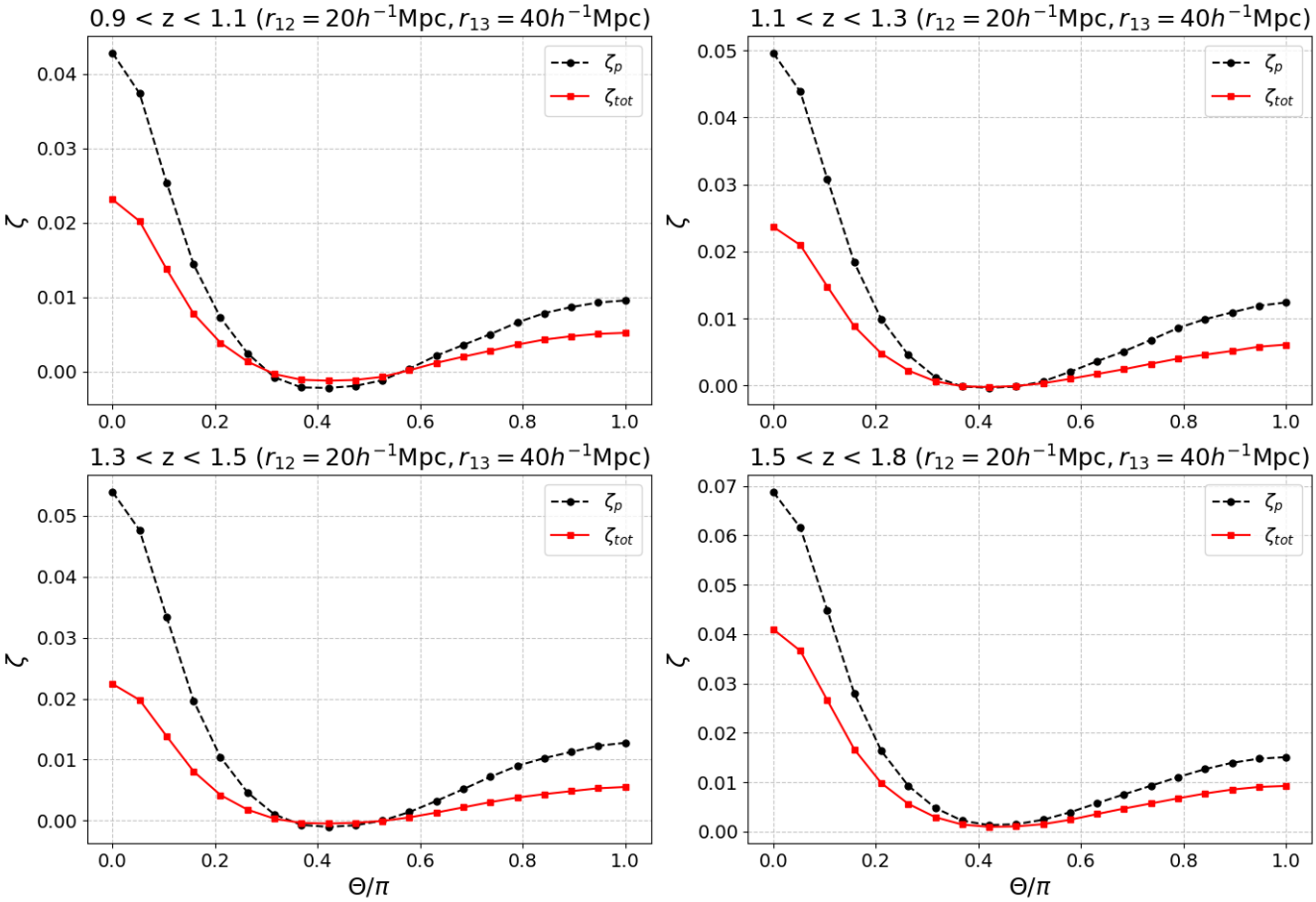


Figure 5.10: *Comparison between the total (black dashed line) and pure (red solid line) signal of the connected 3PCF in the fixed sides configuration ($r_{12} = 20Mpc/h$ and $r_{13} = 40Mpc/h$). Each panel represents a different redhsift bin:* $0.9 < z < 1.1$ *(upper left),* $1.1 < z < 1.3$ *(upper right),* $1.3 < z < 1.5$ *(lower left) and* $1.5 < z < 1.8$ *(lower right).*

across all redshift ranges, with an average reduction of 50%, 55%, 55%, and 35%, respectively. This indicates a significant damping effect on the signal by interlopers, with a slight weakening for $z > 1.5$.

However, the all-scales 3PCF is not easily interpretable. In order to better visualize the effect of the interlopers, we will analyze it for a specific triangular configuration. Figure 5.10 presents a comparison between the connected 3PCF derived from the total FS2 catalogue and the auto-3PCF of the pure sub-catalogue for the configuration $r_{12} = 20\,h^{-1}$Mpc and $r_{13} = 40\,h^{-1}$Mpc. This intermediate scale was chosen because it is optimal for constraining bias parameters—a task addressed in Chapter 6. These scales are large enough to avoid the inaccuracies in models for $r < 10\,h^{-1}$Mpc due to

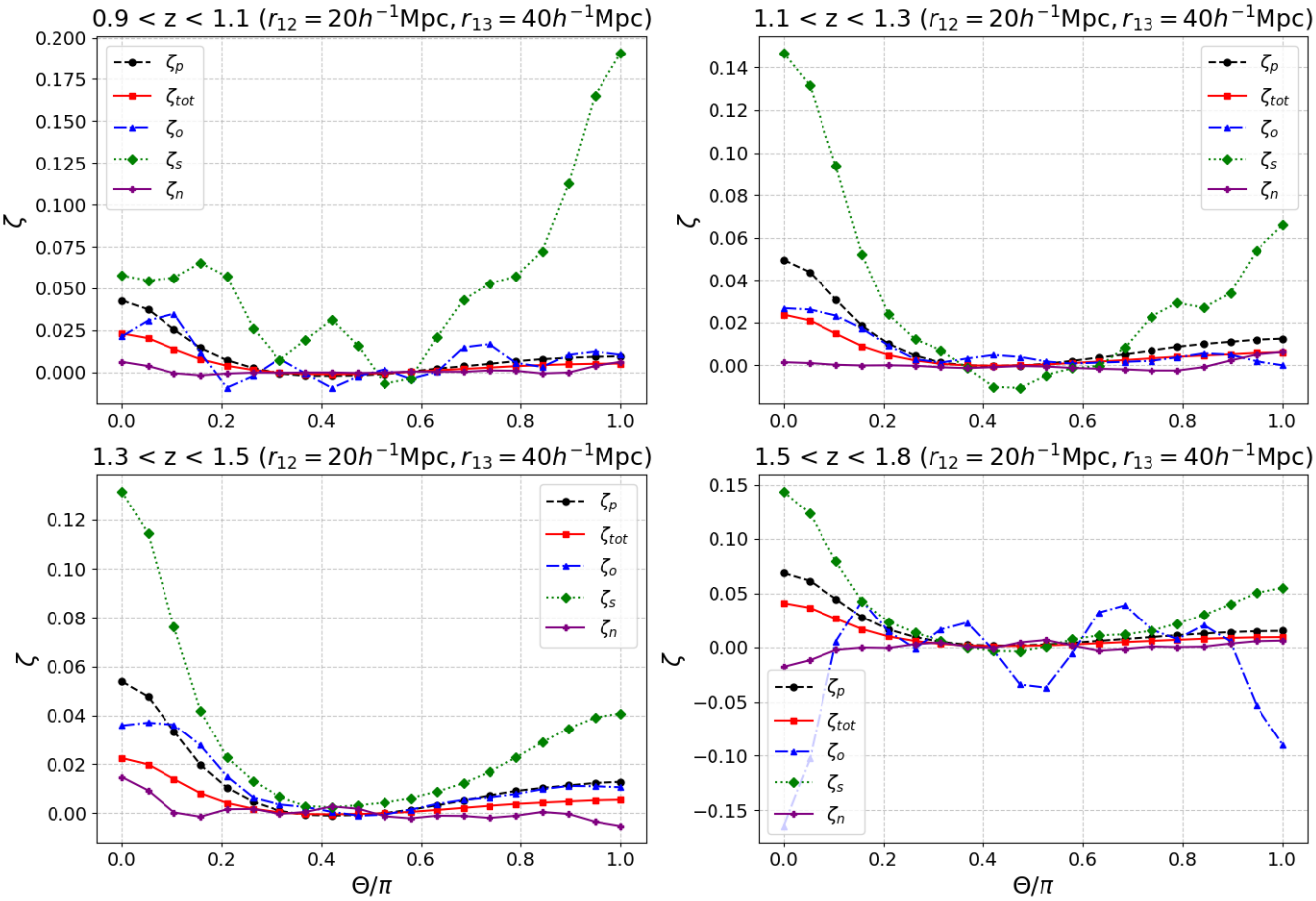


Figure 5.11: *Comparison between the total (black dashed line), pure (red solid line), OIII interlopers (blue dash-dotted line), SIII interlopers (green dotted line), and noise interlopers (purple solid line) signal of the connected 3PCF in the fixed sides configuration ($r_{12} = 20Mpc/h$ and $r_{13} = 40Mpc/h$). Each panel represents a different redhsift bin:* $0.9 < z < 1.1$ *(upper left),* $1.1 < z < 1.3$ *(upper right),* $1.3 < z < 1.5$ *(lower left) and* $1.5 < z < 1.8$ *(lower right).*

nonlinear dynamics, and sufficiently small to remain unaffected by the BAO model considered. Furthermore, we wanted to avoid equilateral configurations, for which we would have needed a much higher number of multipoles for the measurements to be accurate. Moresco et al. (2021) [17] also adopted a similar choice for the same reasons. The overall behavior of the 3PCF is similar to that observed in the two-point case: the total signal is damped due to interloper contamination across all redshift ranges. Table 5.3 shows the average damping of the connected 3PCF signal due to interloper contamination for small $(0.00 - 0.35)$, intermediate $(0.35 - 0.50)$ and large $(0.55 - 1.00)$ values of the angular

	Average $\frac{\zeta_p - \zeta_{tot}}{\zeta_p}$ (small θ/π)	Average $\frac{\zeta_p - \zeta_{tot}}{\zeta_p}$ (intermediate θ/π)	Average $\frac{\zeta_p - \zeta_{tot}}{\zeta_p}$ (large θ/π)
$0.9 < z < 1.1$	47.6%	45.2%	45.6%
$1.1 < z < 1.3$	51.6%	35.4%	52.3%
$1.3 < z < 1.5$	60.6%	50.1%	68.2%
$1.5 < z < 1.8$	40.1%	38.8%	31.2%

Table 5.3: *Average percentual damping of the connected 3PCF coming from the total FS2 catalogue, ζ_{tot}, with respect to the one coming from the pure sub-catalogue, ζ_p. The values are reported for small ($0.00 - 0.35$), intermediate ($0.35 - 0.50$) and large ($0.55 - 1.00$) values of the angular variable θ in units of π, in the first, second and third column respectively. Each row refers to a different redshift range.*

variable, θ, in units of π across the four redshift ranges. The values exceed 30%, touching peaks above 60%, although they remain more or less constant for different scales of θ/π. Once again, we notice that the attenuation is lower in the fourth redshift range, corroborating the hypothesis that the impact of the interlopers is weakened above redshift $z = 1.5$. However, we can conclude that also the 3PCF signal is strongly affected by interloper contamination in all redshift ranges.

Moreover, since the amplitude of the 3PCF signal is closely related to the bias parameters, particularly the linear bias b_1, based on the results in Table 5.3 we expect an error in the estimation of the latter when modelling the total 3PCF. This prediction will be confirmed, as we will show in Chapter 6.

Figure 5.11 illustrates the auto-correlations of the connected 3PCF for the same configuration, this time including the interlopers' auto-3PCFs. However, the latter's behavior is influenced by the small sample sizes of some interloper categories. Three-point statistics require large samples to achieve accuracy, which is not the case e.g. for the catalogues of OIII interlopers in the first and fourth redshift ranges and SIII interlopers in the first redshift range (see Table 5.1). In general, the auto-correlations of the interlopers exhibit similar behavior to their two-point counterparts: the SIII signal is higher than the pure signal, whereas the opposite is true for the OIII signal (for the same reasons explained in Section 5.2). Moreover, the noise signal is consistently zero, as expected for a noise population. Since the uncertainties have not been estimated for these measurements, we will analyze the residuals in the next Paragraph.

5.3.1 Cross-correlations of the three-point correlation function

Following the analysis of the auto-3PCF signals, we moved to the cross-correlation measurements. Due to computational limitations, cross-3PCF measurements were performed in redshift slices with a width of $\Delta z = 0.05$, centered at $z = 1.0, 1.2, 1.4, 1.65$, within each of the four specified redshift intervals. The random and data catalogues were selected

using the same criteria as in the cross-2PCF analysis.

For the cross-3PCF measurements, we employed our new cross-3PCF estimator, as detailed in Section 4.3. For computational time reasons, these measurements were restricted to the fixed-sides configuration $r_{12} = 20\,h^{-1}$Mpc, $r_{13} = 40\,h^{-1}$Mpc with a bin width of $\Delta r = 10\,h^{-1}$Mpc, rather than spanning all comoving scales in the range $15-165\,h^{-1}$Mpc. A random-data ratio of $D_R = 5$ was used. The resulting auto- and cross-3PCF signals for each redshift slice are presented in Figure 5.12. The auto-signals are similar to that described in the preceding Paragraph across all redshift ranges, but with larger Poissonian errors due to smaller catalogue sizes.

Conversely, the cross-signals lack distinct features. Their behaviour is comparable to the auto-noise signal. Notably, these signals exhibit a preference for small values of the angular variable. What we found is consistent with expectations, as there is no compelling reason for galaxies from different sub-catalogues to exhibit strong correlations. The fourth redshift bin is characterized by large error bars for all correlations involving the OIII catalogue, due to the very low percentage of OIII interlopers in this redshift range (see Table 5.1).

The residuals of the total 3PCF, including both auto- and cross-correlations, can be expressed as:

$$\begin{aligned}
\zeta_{\rm res} =& \zeta_{\rm tot} - f_p^3 \frac{R_p R_p R_p}{RRR}\zeta_{ppp} - f_o^3 \frac{R_o R_o R_o}{RRR}\zeta_{ooo} - f_s^3 \frac{R_s R_s R_s}{RRR}\zeta_{sss} - f_n^3 \frac{R_n R_n R_n}{RRR}\zeta_{nnn} \\
&- 3\left(f_p^2 f_o \frac{R_p R_p R_o}{RRR}\zeta_{ppo} + f_p f_o^2 \frac{R_p R_o R_o}{RRR}\zeta_{poo}\right) \\
&- 3\left(f_p^2 f_s \frac{R_p R_p R_s}{RRR}\zeta_{pps} + f_p f_s^2 \frac{R_p R_s R_s}{RRR}\zeta_{pss}\right) \\
&- 3\left(f_p^2 f_n \frac{R_p R_p R_n}{RRR}\zeta_{ppn} + f_p f_n^2 \frac{R_p R_n R_n}{RRR}\zeta_{pnn}\right) \\
&- 3\left(f_o^2 f_s \frac{R_o R_o R_s}{RRR}\zeta_{oos} + f_o f_s^2 \frac{R_o R_s R_s}{RRR}\zeta_{oss}\right) \\
&- 3\left(f_o^2 f_n \frac{R_o R_o R_n}{RRR}\zeta_{oon} + f_o f_n^2 \frac{R_o R_n R_n}{RRR}\zeta_{onn}\right) \\
&- 3\left(f_s^2 f_n \frac{R_s R_s R_n}{RRR}\zeta_{ssn} + f_s f_n^2 \frac{R_s R_n R_n}{RRR}\zeta_{snn}\right) \\
&- 6 f_p f_o f_s \frac{R_p R_o R_s}{RRR}\zeta_{pos} - 6 f_p f_o f_n \frac{R_p R_o R_n}{RRR}\zeta_{pon} \\
&- 6 f_p f_s f_n \frac{R_p R_s R_n}{RRR}\zeta_{psn} - 6 f_o f_s f_n \frac{R_o R_s R_n}{RRR}\zeta_{osn}.
\end{aligned} \tag{5.9}$$

As illustrated in Figure 5.13, the residuals consistently approach zero across all four redshift slices, thereby validating Eq.(3.33).

This represents the first measurement within the *Euclid* collaboration of the impact of redshift interlopers on the measurement of the 3PCF. It constitutes a pivotal step towards the accurate utilization of this tool in cosmology.

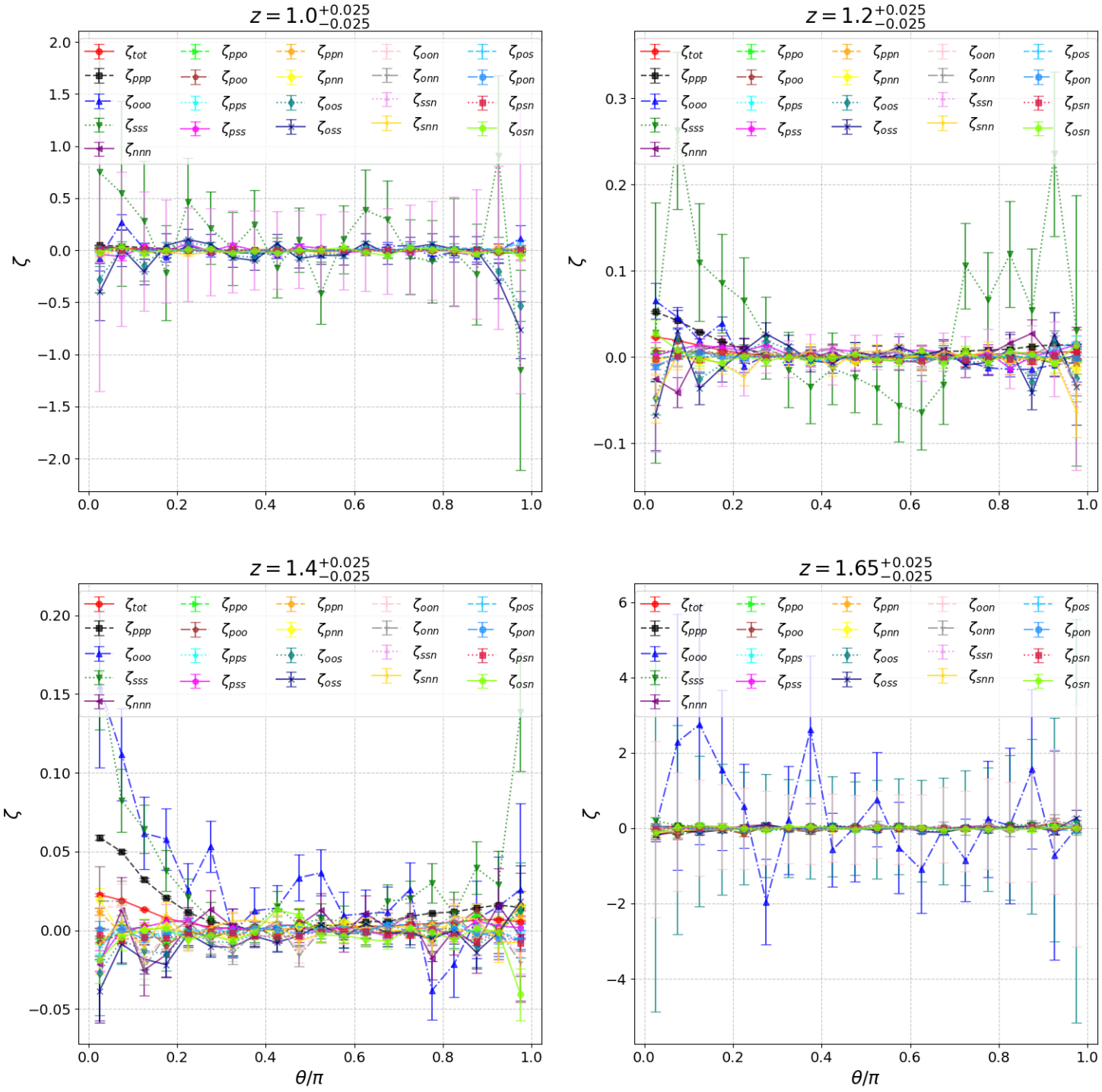


Figure 5.12: *Comparison between the total connected 3PCF signals (solid red line), the auto-3PCF of pure galaxies (black dashed line), OIII interlopers (blue dash-dotted line), SIII interlopers (green dotted line), and noise interlopers (purple solid line), and all the other cross-3PCF signals in the fixed sides configuration ($r_{12} = 25 Mpc/h$ and $r_{13} = 45 Mpc/h$). Each panel represents a different redshift range. Error bars represent Poissonian uncertainties.*

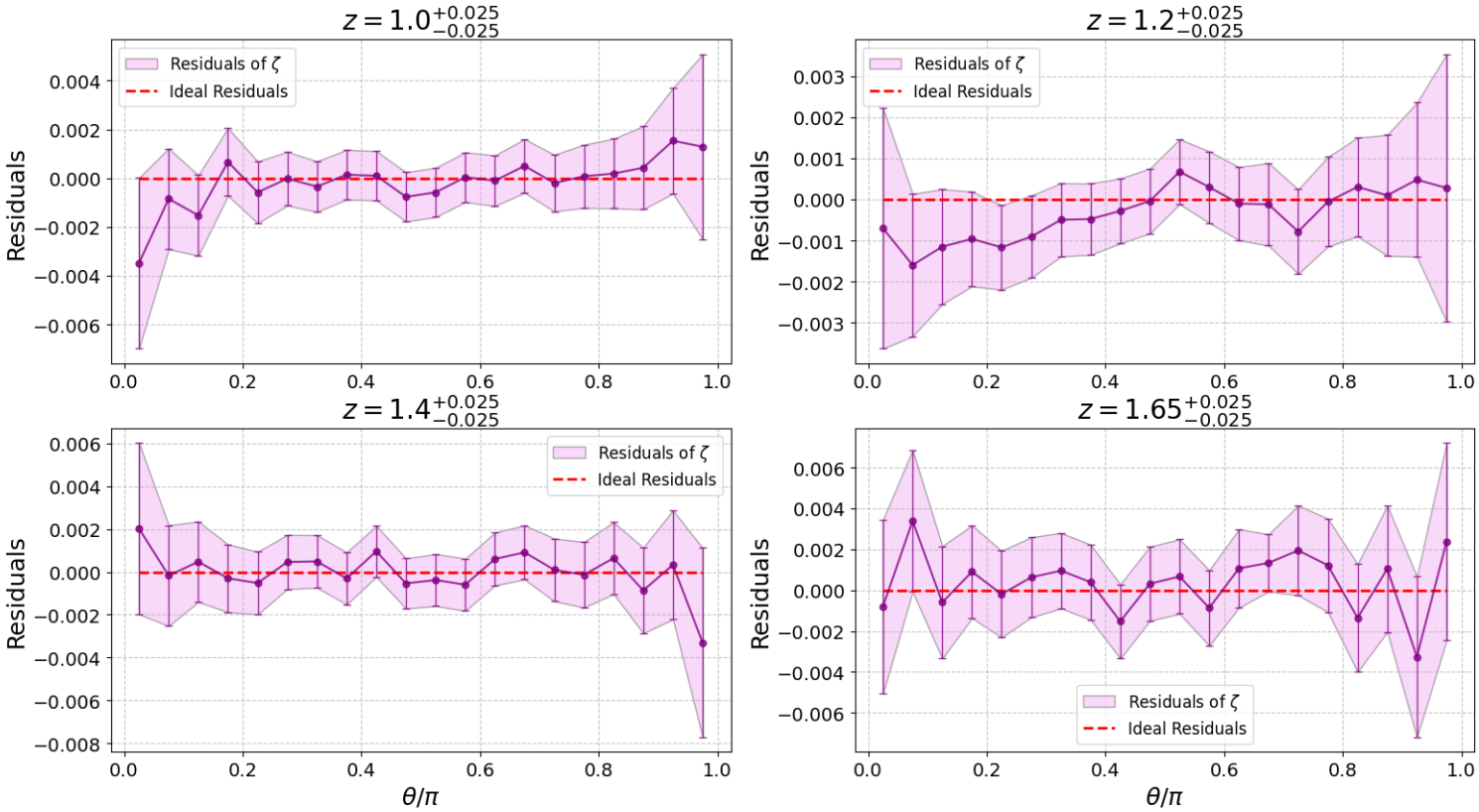


Figure 5.13: *Residuals of the connected 3PCF signal of the FS2 catalogue and all the signals coming from the combination of its sub-catalogues (auto-signals and cross-signals). The residuals are computed using Eq.*(5.7)*. The ideal residuals (zero) are indicated with a red dashed line. Each panel represents a different redshift range. Error bars represent Poissonian uncertainties.*

Chapter 6

Measurements and modelling from the EuclidLargeMocks

Following our analysis of the effects of interloper contamination on the measurement of the 2PCF and 3PCF, we directed our attention to the final and most compelling aspect of our study: the impact of interloper contamination on the constraints of cosmological parameters derived through the modelling of 2PCF and 3PCF. For this, we used the EuclidLargeMocks, which will be detailed in the subsequent Section. We will then present our measurements for both 2PCF and 3PCF, alongside the estimated covariance matrices, which are essential for accurate modelling. Finally, we will discuss the constraints obtained on the cosmological parameters and evaluate the influence of interloper contamination on these constraints.

6.1 EuclidLargeMocks

The EuclidLargeMocks comprise a set of 1000 mock catalogues generated using simulations performed with *PINOCCHIO v5.0*. This code exploits the Lagrangian Perturbation Theory to generate catalogues of cosmological dark matter halos, providing information on their mass, position, velocity, and merger history (Munari, Monaco et al., 2016 [48]). Each mock catalogue represents a 30-degree field on the sky plane, with the majority of galaxies within the redshift range $0.9 < z < 1.8$.

For each mock, a corresponding mask-catalogue is available, which introduces redshift systematic errors using the method employed by de la Torre et al. (in prep.) for the FS2 catalogue, as discussed in the previous Chapter. By combining each mock with its mask, we can generate interloper-contaminated catalogues where galaxies are flagged by type. This enables us to process the mocks similarly to the FS2 catalogue, dividing them into mutually exclusive sub-catalogues of homogeneous galaxy types.

Figure 6.1 presents the total galaxy redshift distribution, $n(z)$, of the first mock and

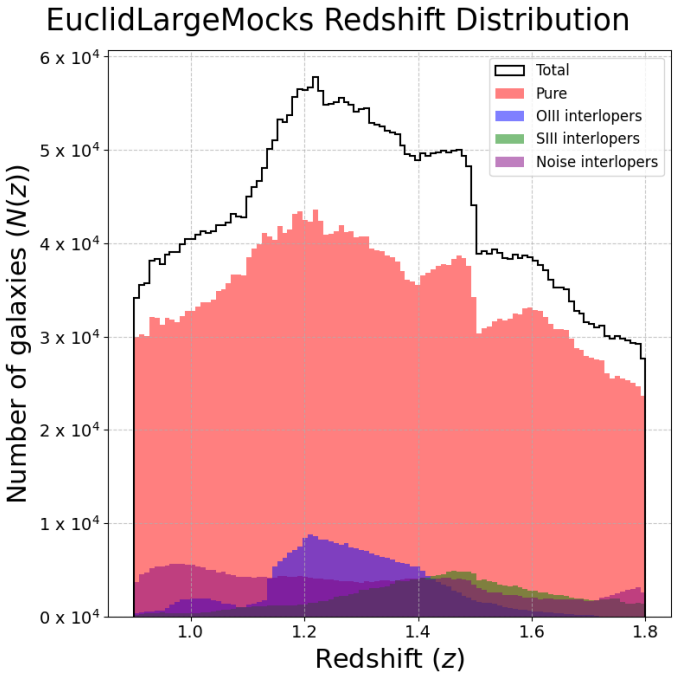


Figure 6.1: *Redshift distribution in the range* $0.9 < z < 1.8$ *of the total first catalogue of the ELM (solid black line) and of its four sub-catalogues: pure, OIII, SIII and noise (red, blue, green and purple, respectively).*

its sub-catalogues (for the rest of the mocks the distributions are almost identical). The similarity to the redshift distributions of FS2 (Figure 5.2) arises from the use of the same algorithm for inserting systematics, with the peculiar shape of $n(z)$ explained in Section 5.1. As with the FS2 catalogue, we divided the mocks into four redshift ranges: $0.9 < z < 1.1$, $1.1 < z < 1.3$, $1.3 < z < 1.5$, and $1.5 < z < 1.8$. The percentage composition of each redshift range is detailed in Table 6.1.

Random catalogues for both the total and sub-catalogues were constructed as in the previous Chapter (Lee et al., in prep. and Risso et al., in prep.). Figure 6.2 compares the galaxy density distributions of the first mock within the first redshift range to the corresponding random catalogue.

6.2 Two- and three-point correlation function measurements

To investigate the potential bias in cosmological parameters due to interloper contamination, we first measured the 2PCF and 3PCF signals for the ELM, focusing on both

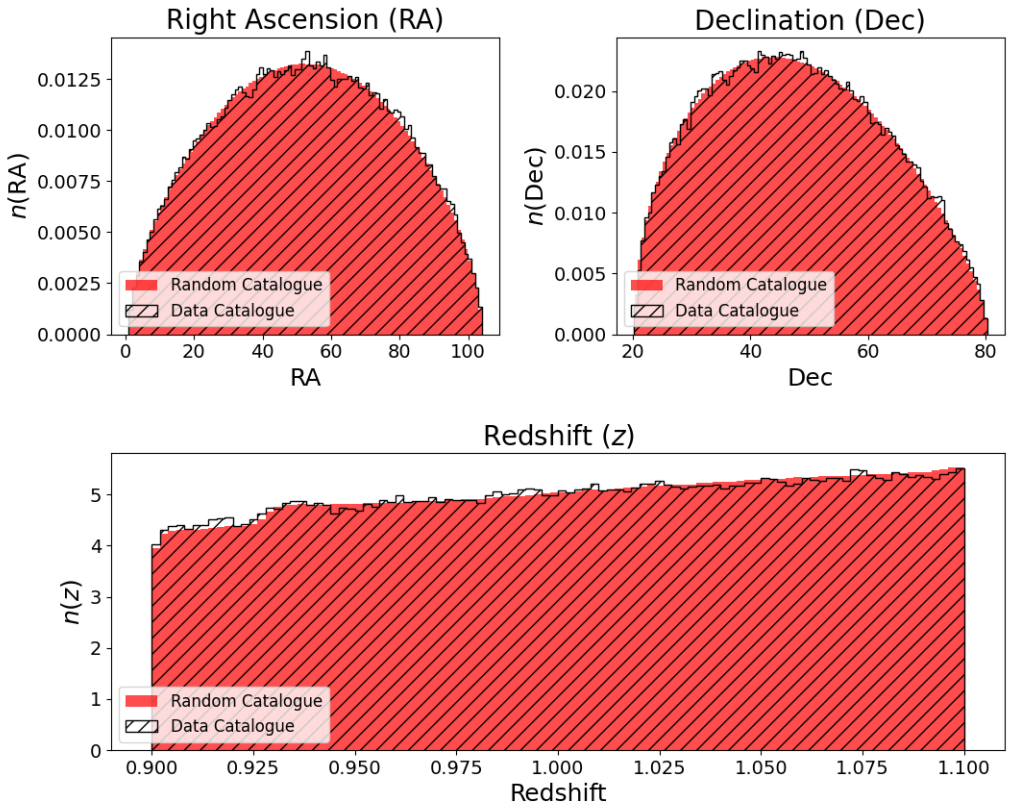


Figure 6.2: *Comparison of the Right Ascension, Declination, and redshift distributions ($n(RA)$, $n(Dec)$, and $n(z)$, respectively) of galaxies in the first ELM catalogue (black line) and the corresponding random catalogue (red) within the first redshift range ($0.1 < z < 1.1$).*

the total signal and the signal derived from the pure sub-catalogue. Specifically, due to computational constraints, we measured the first three even multipoles of the 2PCF and the connected 3PCF for the initial 200 mocks within the redshift ranges $0.9 < z < 1.1$ and $1.1 < z < 1.3$.

The results of the 2PCF measurements are presented in Figure 6.3. These measurements were obtained from the mocks within a comoving separation range of $0 - 180\,h^{-1}$Mpc, with a linear binning interval of $\Delta r = 5\,h^{-1}$Mpc. To estimate the multipoles of the 2PCF we used a CBL function which exploits the direct LS estimator (Eq. 3.7). A random-data ratio of $D_R = 25$ was used. Figure 6.3 illustrates both the individual mock signals and the average signal, indicating that the total 2PCF signal is damped

	Pure galaxies	OIII interlopers	SIII interlopers	Noise interlopers
$0.9 < z < 1.1$	83.1%	3.2%	1.2%	12.5%
$1.1 < z < 1.3$	77.6%	11.8%	2.9%	7.7%
$1.3 < z < 1.5$	74.9%	9.4%	7.9%	7.8%
$1.5 < z < 1.8$	85.0%	1.4%	7.1%	6.5%

Table 6.1: *Percentage composition of the first catalogue of the ELM in each redshift range. Rows show the redshift range, while columns distinguish between different sub-catalogues.*

compared to the pure signal across all three multipoles (monopole, ξ_0, quadrupole, ξ_2, and hexadecapole, ξ_4), corroborating our findings from the FS2 catalogue. The displayed errors have been derived from the diagonal of the covariance matrix. Furthermore, in Table 6.2 we provide the average damping of the total 2PCF monopole with respect to the signal from the pure sub-catalogues $(\xi_{0,p} - \xi_{0,tot})/\xi_{0,p}$. The values confirm the study we carried out using the FS2 catalogue, indicating an attenuation of the signal around 30% at small scales for both redshift ranges. Moreover, also in this case the results indicate a very strong damping at large scales for the first redshift range ($\sim 70\%$), while for $1.1 < z < 1.3$ the attenuation is almost constant with respect to the scale.

Figure 6.4 provides an example of the 3PCF measurement results. These measurements were conducted using the same technique as for the FS2 catalogue. The auto-3PCF multipoles were measured up to $\ell = 11$ for all fixed side configurations with r_{12} and r_{13} spacing in the range of $20 - 130\,h^{-1}$Mpc, with a binning interval of $\Delta r = 10\,h^{-1}$Mpc, utilizing the CBL function which exploits the SHD method proposed by Slepian and Eisenstein (2015) [40]. A random-data ratio of $D_R = 25$ was used. Subsequently, the multipoles were combined to reconstruct the connected 3PCF signal for two fixed side configurations within the $20 - 130\,h^{-1}$Mpc range, using 20 bins for the angular variable. Figure 6.4 confirms the results from the FS2 catalogue, showing a damped total signal relative to the pure signal in the connected 3PCF for the fixed sides $r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 45\,h^{-1}$Mpc. As with the 2PCF results, the plot displays both the individual mock signals and the mean signal, with errors derived from the diagonal of the covariance matrix. Table 6.3 shows the average damping of the connected 3PCF signal of the total catalogues with respect to the one coming from the pure sub-catalogues, $(\zeta_p - \zeta_{tot})/\zeta_p$. The values indicate a constant attenuation for low, intermediate and large values of the angular variable, θ, that exceeds 40% and 50% for the first and second redshift range, respectively. As the Table shows, the damping is dramatic and its effect on the cosmological parameter constraints must be investigated. This task will be addressed in the next Section.

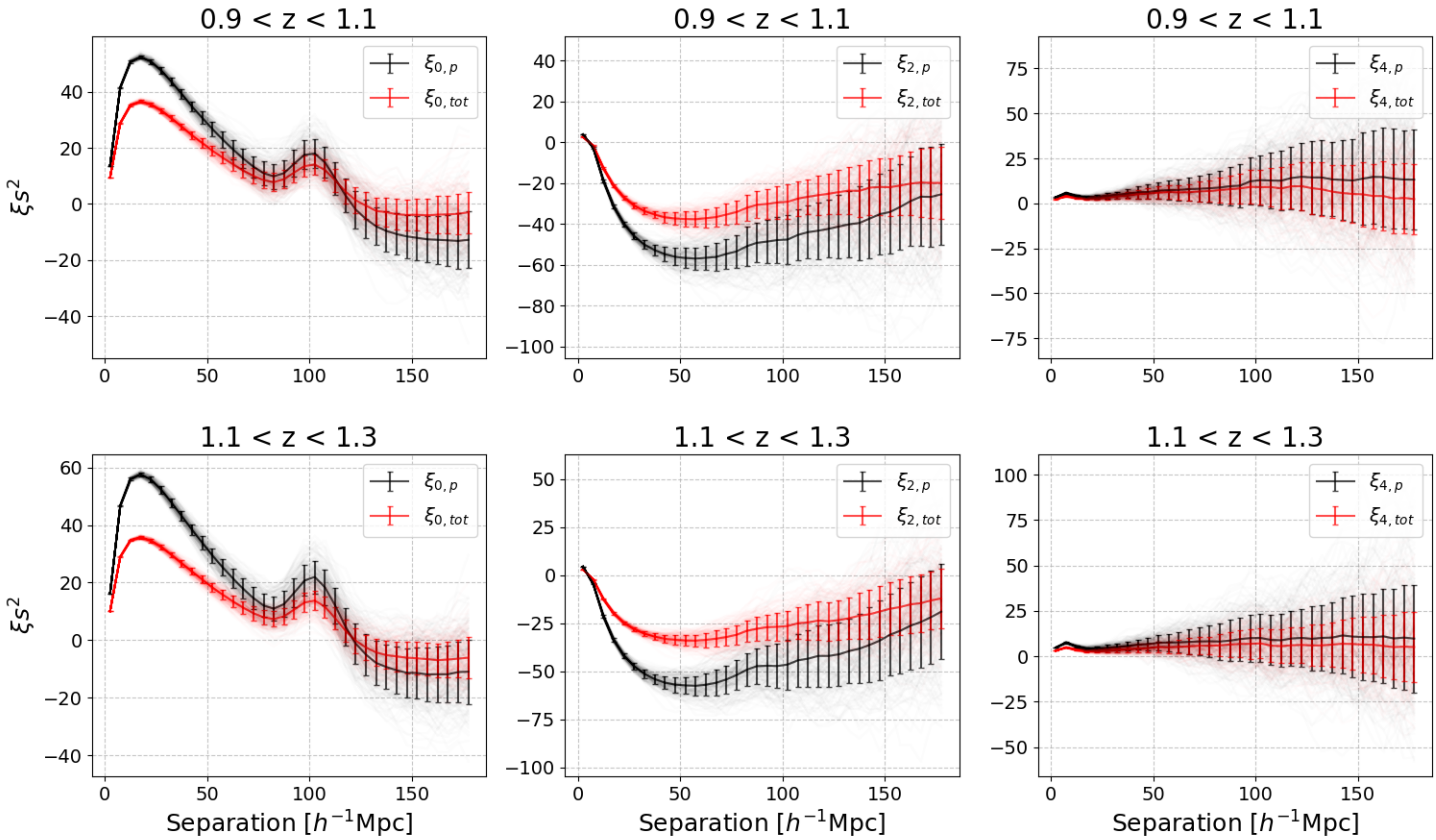


Figure 6.3: *Comparison between the mean total 2PCF signals (solid red line) and the mean auto-2PCF of pure galaxies (solid black line) of the first 200 ELM, within a co-moving separation range of* $0 - 180h^{-1}$ *Mpc. The upper panels show the first three even multipoles of the signal (monopole, quadrupole, and hexadecapole from left to right) for the first redshift bin* ($0.9 < z < 1.1$)*, and the lower panels show the signal for the second redshift bin* ($1.1 < z < 1.3$)*. Error bars are estimated from the diagonal of the covariance matrix. The less visible lines represent the measurement for each mock total catalogue (red lines) and pure sub-catalogue (black lines).*

6.2.1 Covariance matrices

For the modelling of the 2PCF and 3PCF signals, the uncertainty in the measurements is required. The latter is quantified by the covariance matrix C_{ij}, defined as follows:

$$C_{ij} \equiv E\left[(x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)\right], \tag{6.1}$$

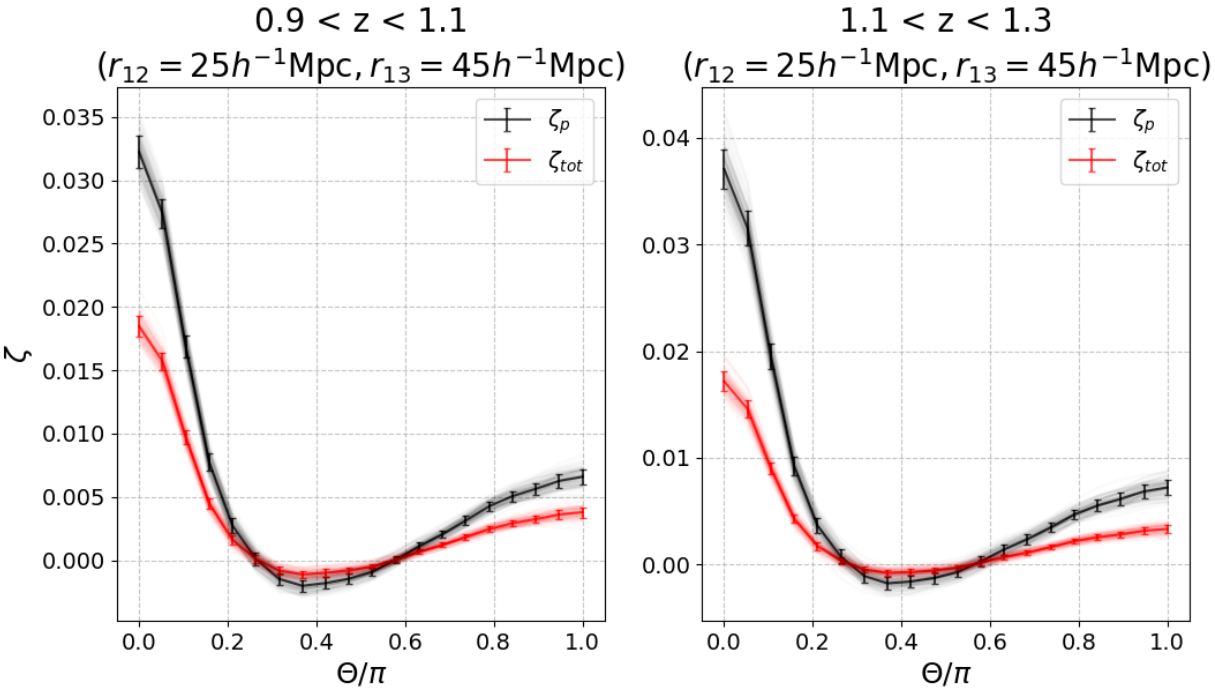


Figure 6.4: *Comparison between the mean total 3PCF signals (solid red line) and the mean auto-3PCF of pure galaxies (solid black line) of the first 200 ELM, for the fixed-sides configurations, with sides* $r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 45\,h^{-1}$Mpc. *The left panel shows the connected 3PCF for the first redshift bin* ($0.9 < z < 1.1$)*, and the right panel shows the signal for the second redshift bin* ($1.1 < z < 1.3$)*. Error bars are estimated from the diagonal of the covariance matrix. The less visible lines represent the measurement for each mock total catalogue (red lines) and pure sub-catalogue (black lines).*

where x_i and x_j are two variables, $E(x)$ represents the expectation value of the variable x, and $\langle x \rangle$ denotes the mean value of the variable x. The covariance matrix for our measurements was computed using a CBL function that employs the following maximum likelihood estimator [49]:

$$\hat{C}_{ij}^{\text{ML}} = \frac{1}{N_m - 1} \sum_{k=1}^{N_m} (d_i^{(k)} - \bar{d}_i)(d_j^{(k)} - \bar{d}_j), \tag{6.2}$$

where N_m is the number of mock catalogues, $d_i^{(k)}$ represents the data in the i-th bin of the k-th mock catalog, and $\bar{d}_i$ is the mean value of the data in the i-th bin, averaged over all mock catalogues. For better visualization, instead of the covariance matrix, we present the correlation matrix of our data, defined as follows:

$$\text{Corr}_{ij} \equiv \frac{C_{ij}}{\sqrt{C_{ii} C_{jj}}}. \tag{6.3}$$

	Average small scale $\frac{\xi_{0,p}-\xi_{0,tot}}{\xi_{0,p}}$	Average large scale $\frac{\xi_{0,p}-\xi_{0,tot}}{\xi_{0,p}}$
$0.9 < z < 1.1$	26.5%	71.1%
$1.1 < z < 1.3$	36.6%	43.4%

Table 6.2: *Average percentual damping of the mean 2PCF monopole signal coming from the first 200 ELM catalogues, $\xi_{0,tot}$, with respect to the one coming from the first 200 pure sub-catalogues, $\xi_{0,p}$. The values are reported for small scales ($0 - 120\,h^{-1}$Mpc) in the first column and for large scales ($150 - 180\,h^{-1}$Mpc) in the second column. Each row refers to a different redshift range.*

	Average $\frac{\zeta_p-\zeta_{tot}}{\zeta_p}$ (small θ/π)	Average $\frac{\zeta_p-\zeta_{tot}}{\zeta_p}$ (intermediate θ/π)	Average $\frac{\zeta_p-\zeta_{tot}}{\zeta_p}$ (large θ/π)
$0.9 < z < 1.1$	42.1%	44.6%	42.5%
$1.1 < z < 1.3$	53.4%	53.9%	55.5%

Table 6.3: *Average percentual damping of the connected 3PCF in fixed sides configuration ($r_{12} = 25\,h^{-1}Mpc$, $r_{13} = 45\,h^{-1}Mpc$) coming from the first 200 ELM catalogues, ζ_{tot}, with respect to the one coming from the first 200 pure sub-catalogues, ζ_p. The values are reported for small ($0.00 - 0.35$), intermediate ($0.35 - 0.50$) and large ($0.55 - 1.00$) values of the angular variable θ in units of π, in the first, second and third column respectively. Each row refers to a different redshift range.*

Figure 6.5 displays the correlation matrix of the 2PCF multipole measurements from the total and pure samples in the first two redshift ranges. The matrix is 108×108 in size, as we have a total of 108 radial bins (36 for each multipole), and it is divided into 9 sub-squares. The three diagonal sub-squares show the correlation between measurements of the same multipole, while the other six show the cross-correlation between different multipoles.

Figure 6.6 presents the correlation matrix corresponding to the measurement of the connected 3PCF in the fixed sides configuration ($r_{12} = 25\,h^{-1}$Mpc, $r_{13} = 45\,h^{-1}$Mpc). The matrix is 20×20 in size, as we have a total of 20 angular bins.

6.3 Two- and three-point correlation function modelling

Bayesian analysis is a common method used to constrain cosmological parameters from measurements. This statistical approach enables probabilistic inference about unknown quantities based on available data and prior knowledge, exploiting the Bayes theorem. The latter states, considering two events, A and B, that:

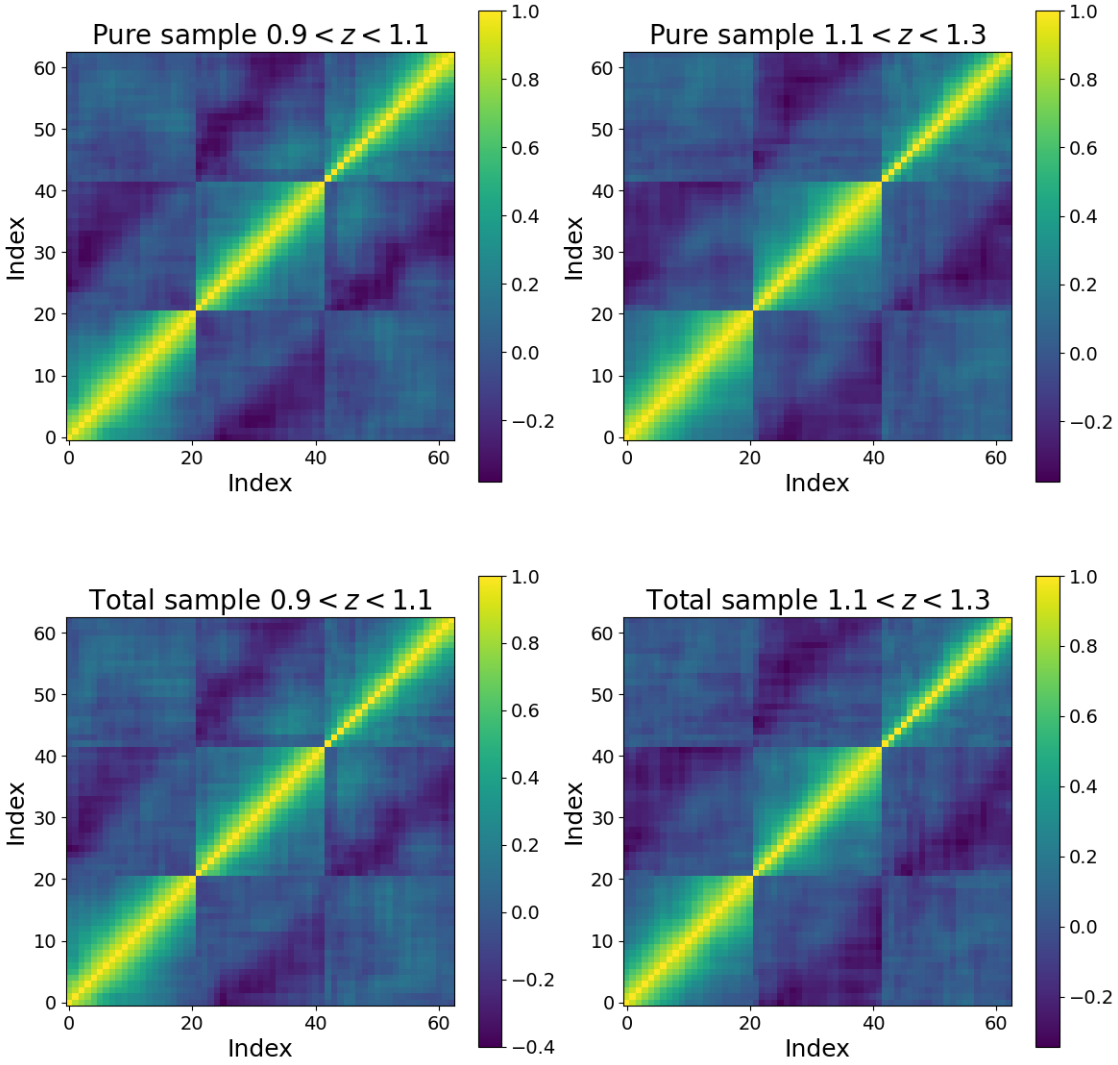


Figure 6.5: *Correlation matrices of the 2PCF multipoles measurements from the first 200 ELM catalogues. The matrices are composed by nine sub-squares. The three diagonal sub-blocks correspond to the 2PCF monopole, quadrupole and hexadecapole, respectively in the bottom-left, central and top-right regions. The other sub-blocks represent the cross-correlation between the different moments. The top panels represent the correlation matrices of the measurements from the pure sub-catalogues (first and second redshift range from left to right), while the bottom panels represent the correlation matrices of the measurement from the total ELM catalogues.*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

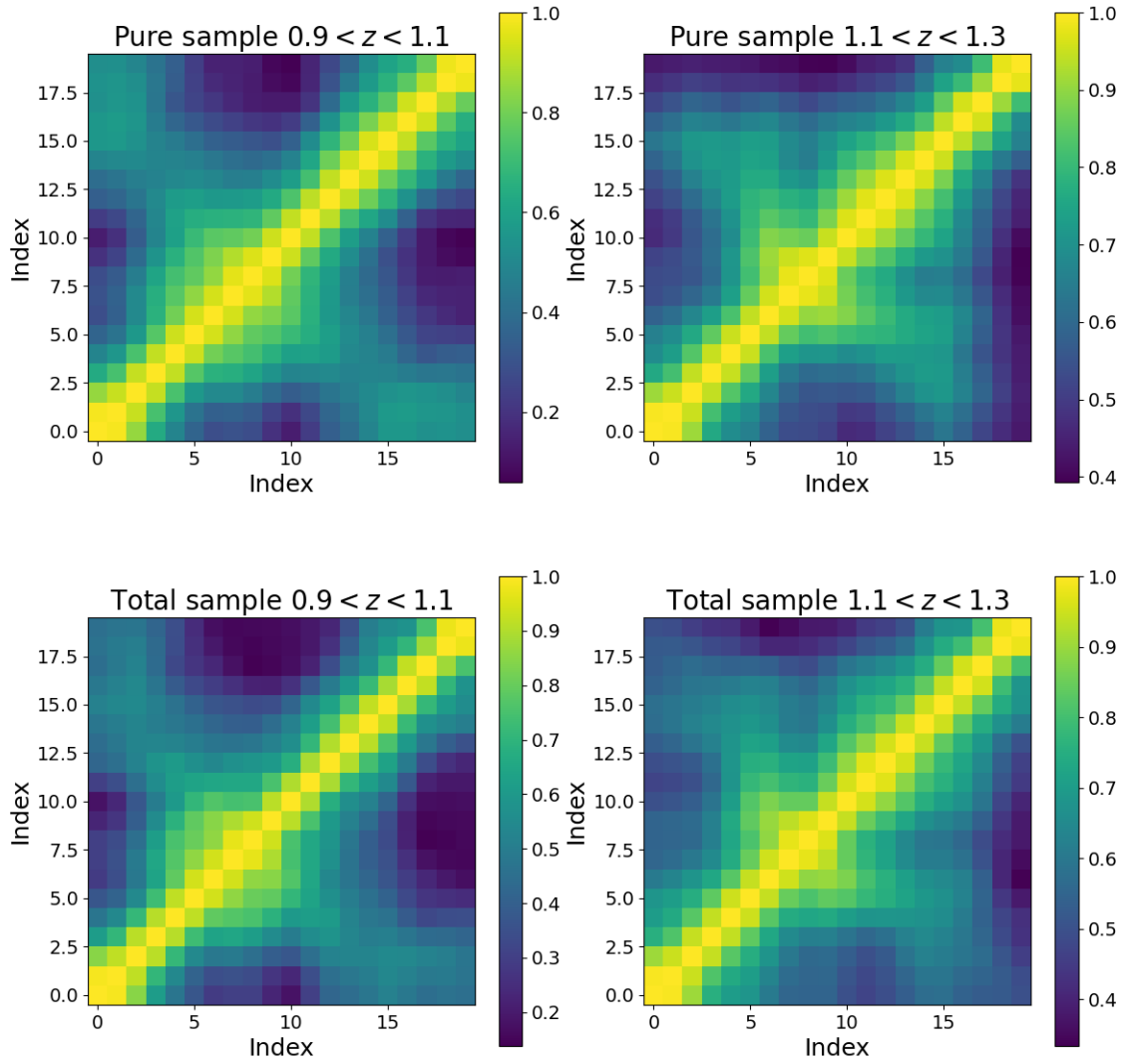


Figure 6.6: *Correlation matrices of the connected 3PCF measurements in the fixed sides configuration ($r_{12} = 25\,h^{-1}Mpc$, $r_{13} = 45\,h^{-1}Mpc$) from the first 200 ELM catalogues. The top panels represent the correlation matrices of the measurements from the pure sub-catalogues (first and second redshift range from left to right), while the bottom panels represent the correlation matrices of the measurement from the total ELM catalogues.*

where $P(A|B)$ is the conditional probability of event A given event B, and $P(B|A)$ is the conditional probability of event B given event A.

In the context of cosmological parameter estimation, let D denote a set of data and $M(\theta)$ represent a model parameterized by the vector θ. The posterior distribution

$f\,\sigma_8$	$b_1\,\sigma_8$	$b_2\,\sigma_8$	σ_v	$\alpha_\perp$	$\alpha_\parallel$
$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$	$\mathcal{U}(-2,2)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$

Table 6.4: *Prior distributions for the parameters used in the ELM 2PCF analysis, where* $\mathcal{U}(a,b)$ *indicates a uniform distribution between* a *and* b*.*

$P(M(\theta)|D)$ is then defined as the conditional probability of obtaining the model $M(\theta)$ given the data D. Applying the Bayes theorem, we have the following relation:

$$P(M(\theta)|D) = \frac{P(D|M(\theta))P(M(\theta))}{P(D)},$$

where $P(D|M(\theta))$, known as the likelihood function $\mathcal{L}(D|M(\theta))$, quantifies the probability of observing our data set D given the model $M(\theta)$. $P(M(\theta))$, or $\pi(M(\theta))$, is the prior distribution representing our initial knowledge about model $M(\theta)$ before considering the data. $P(D)$ is the evidence, which can be expressed in the following way:

$$P(D) = \int \mathcal{L}(D|M(\theta))\pi(M(\theta))\,\mathrm{d}\theta.$$

Since $P(D)$ is independent of the model parameters θ, it serves as a normalization constant. Thus, the posterior distribution expression simplifies to:

$$P(M(\theta)|D) \propto \mathcal{L}(D|M(\theta))\pi(M(\theta)),$$

where the evidence $P(D)$ is omitted as it acts only as a scaling factor and does not affect the relative probabilities of different models $M(\theta)$. In this work, we used a multi-variate Gaussian likelihood, which can be written as:

$$\mathcal{L} = \frac{1}{(2\pi)^{N_b/2}|C|^{1/2}} e^{-\frac{1}{2}\sum_{ij}[d_i-\mu_i(\theta)]C_{ij}^{-1}[d_j-\mu_j(\theta)]}, \tag{6.4}$$

where $|C| \equiv \mathrm{det}C_{ij}$, d_i is the data at the i-th bin, and $\mu_i(\theta)$ is the model prediction at the i-th bin. Our objective is to analyze the posterior distributions of the parameters, which can be retrieved by multiplying the priors for the likelihood written in Eq.(6.4). We carried out this procedure for both the 2PCF and 3PCF measurements (using the models described in Chapter 3) from the total ELM catalogues and their pure sub-catalogues, in the first two redshift ranges. In this way, we could constrain the bias on the cosmological parametes retrieved by the study of the total samples (which suffers from interlopers contamination) with respect to the ones retrieved by the study of the pure samples (i.e. the parameters unaffected by systematic uncertainties).

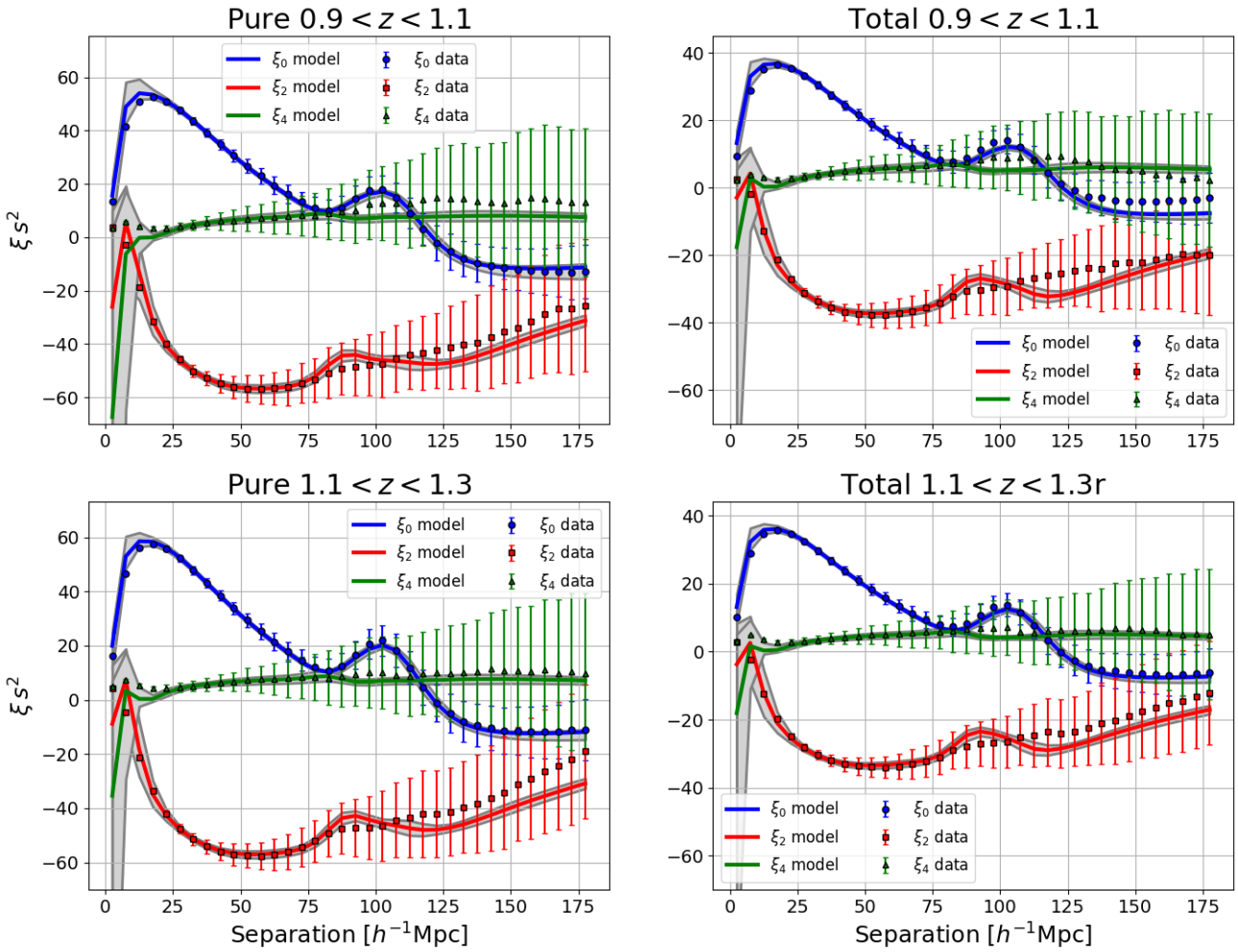


Figure 6.7: *Multipoles (monopole in blue, quadrupole in red, hexadecapole in green) of the total (right panels) and pure (left panels) ELM 2PCF in the first (upper panels) and second (lower panels) redshift range. Circles, squares and triangles represent the data. The error bars are the square root of the diagonal elements of the covariance matrix. The solid curves show the best-fit model, while the grey shaded areas represent the 68% confidence region of the model.*

6.3.1 Two-point correlation function modelling

We started our analysis with the 2PCF modeling, fitting the mean signal from the ELM using the eTNS model, described in Section 3.2. The posterior distribution was sampled using a Monte Carlo Markov Chain (MCMC) with 64 walkers, with a lenght of 3000. The model parameters included $f\,\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, σ_v, $\alpha_\perp$, and $\alpha_\parallel$, where f denotes the growth rate of cosmic structures, σ_8 represents the amplitude of matter density fluctuations, b_1 is the linear bias, b_2 is the nonlinear bias, σ_v is the velocity dispersion, and $\alpha_\perp$ and $\alpha_\parallel$ are parameters related to the AP effect as defined in Section 2.3.2. All parameters were

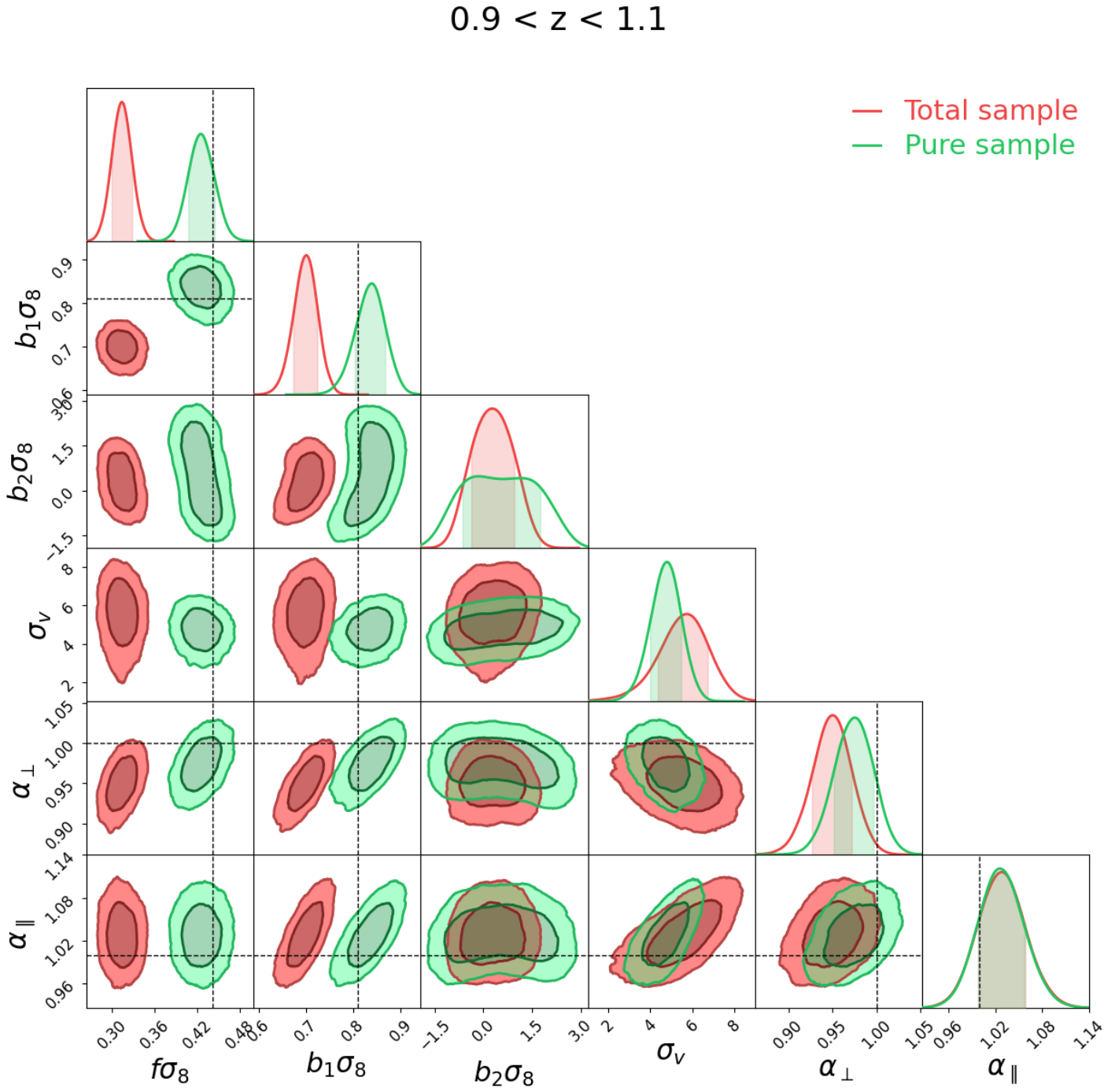


Figure 6.8: *1D and 2D posterior distributions for the free parameters $f\,\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, σ_v, $\alpha_\perp$ and $\alpha_\parallel$ of the eTNS model, obtained from the analysis of the total (in red) and pure (in green) ELM 2PCF multipoles in the redshift range $0.9 < z < 1.1$. The darker and lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

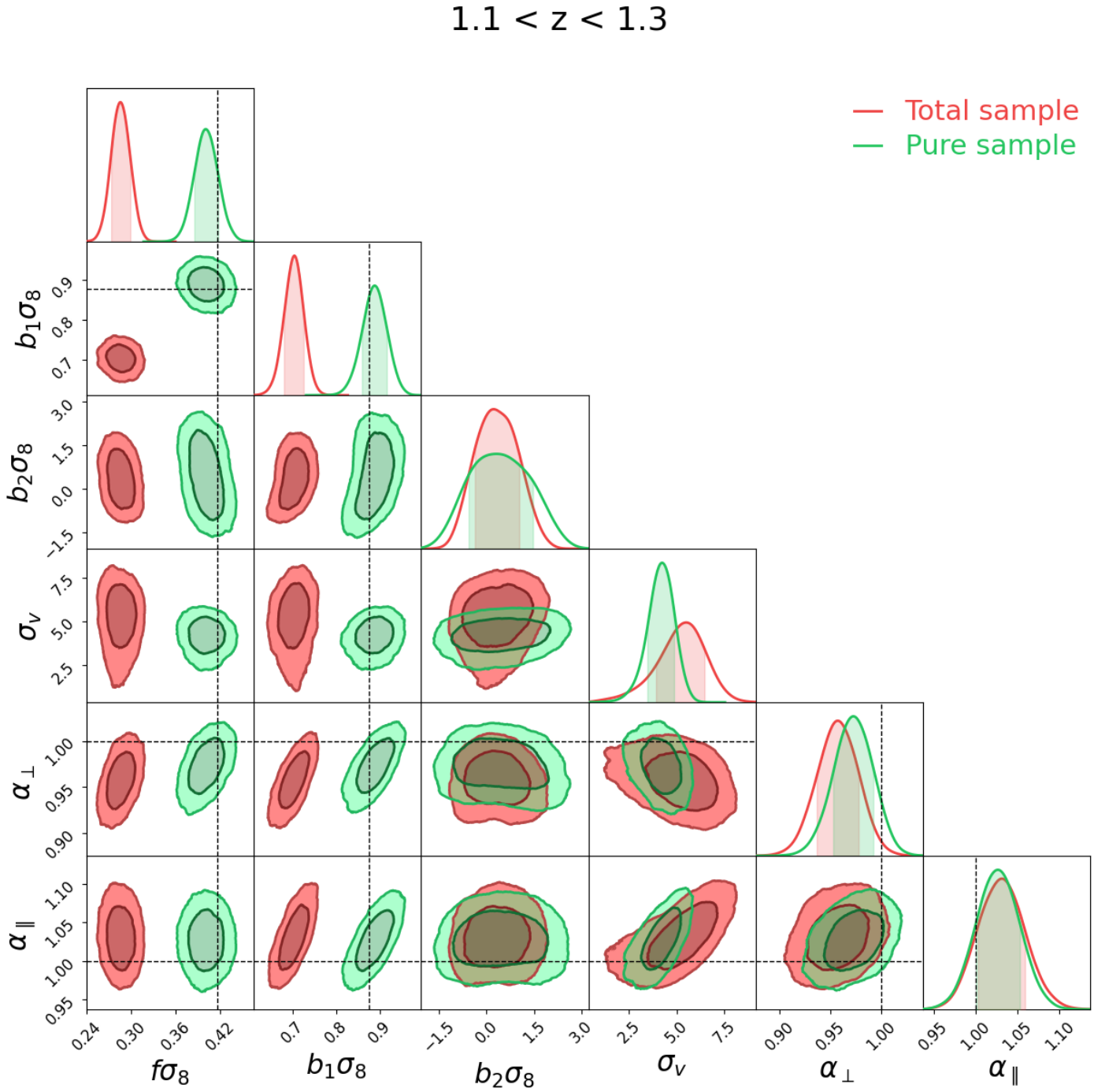


Figure 6.9: *1D and 2D posterior distributions for the free parameters* $f\,\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, σ_v, $\alpha_\perp$ *and* $\alpha_\parallel$ *of the eTNS model, obtained from the analysis of the total (in red) and pure (in green) ELM 2PCF multipoles in the redshift range* $1.1 < z < 1.3$. *The darker and lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

Model	$f\,\sigma_8$	$b_1\,\sigma_8$	$b_2\,\sigma_8$	σ_v	$\alpha_\perp$	$\alpha_\parallel$
Total sample	0.314 ± 0.014	$0.700^{+0.024}_{-0.025}$	$0.28^{+0.65}_{-0.64}$	$5.6^{+1.1}_{-1.2}$	0.950 ± 0.022	$1.028^{+0.030}_{-0.029}$
Pure sample	0.426 ± 0.018	$0.837^{+0.030}_{-0.032}$	0.5 ± 1.2	$4.75^{+0.69}_{-0.74}$	0.974 ± 0.022	$1.027^{+0.029}_{-0.028}$

Table 6.5: *Best-fit values and 68% confidence region for the free parameters $f\,\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, σ_v, $\alpha_\perp$ and $\alpha_\parallel$ of the eTNS model, obtained from the analysis of the total (first row) and pure (second row) ELM 2PCF multipoles in the redshift range* $0.9 < z < 1.1$.

Model	$f\,\sigma_8$	$b_1\,\sigma_8$	$b_2\,\sigma_8$	σ_v	$\alpha_\perp$	$\alpha_\parallel$
Total sample	$0.286^{+0.013}_{-0.012}$	$0.704^{+0.021}_{-0.022}$	$0.32^{+0.70}_{-0.67}$	$5.3^{+1.1}_{-1.4}$	0.957 ± 0.020	1.031 ± 0.027
Pure sample	$0.401^{+0.016}_{-0.015}$	$0.888^{+0.027}_{-0.028}$	$0.43^{+1.03}_{-0.96}$	$4.20^{+0.65}_{-0.71}$	0.973 ± 0.019	1.027 ± 0.025

Table 6.6: *Best-fit values and 68% confidence region for the free parameters $f\,\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, σ_v, $\alpha_\perp$ and $\alpha_\parallel$ of the eTNS model, obtained from the analysis of the total (first row) and pure (second row) ELM 2PCF multipoles in the redshift range* $1.1 < z < 1.3$.

allowed to vary freely with uniform prior distributions, as listed in Table 6.4.

Figure 6.7 displays the data alongside the best-fit model for both the total and pure samples across two redshift ranges. The fitting range was set between $25 - 130\,h^{-1}$Mpc, within which the model accurately reproduced the monopole of the 2PCF, ξ_0. Some discrepancies can be noted in the total monopole within the first redshift range, although these occurred outside the fitting range. Additionally, the 68% confidence region error at small scales was notably broad, attributable to current models' challenges in handling nonlinear scales (below $\sim 20\,h^{-1}$Mpc) due to baryonic effects.

Figures 6.8 and 6.9 present the 1D and 2D posterior distributions for the total and pure fits in the first and second redshift ranges, respectively. The best-fit values and 68% confidence regions are summarized in Tables 6.5 and 6.6. The posterior distributions of the total samples exhibited a significant offset compared to the pure samples. Specifically, $f\,\sigma_8$ was attenuated by 26.3% and 28.7% in the first and second redshift ranges, respectively, while $b_1\,\sigma_8$ was reduced by 16.4% and 20.7%. Conversely, σ_v posteriors showed an increase in both width and best-fit value by 17.9% and 26.2% in the first and second redshift ranges, respectively. Interestingly, the $b_2\,\sigma_8$ parameter has compatible best-fit values between the total and pure samples, with the posterior distribution indicating improvement for the total sample. Regarding the AP parameters, the $\alpha_\parallel$ posterior distributions appeared unaffected by interloper contamination, whereas $\alpha_\perp$ experienced a damping of 2.5% and 1.6% in the first and second redshift ranges, respectively.

These findings indicate that interloper contamination significantly impacts the constraints on cosmological parameters, particularly $f\,\sigma_8$ and $b_1\,\sigma_8$. The notable behavior of $b_2\,\sigma_8$ suggests an intriguing feature warranting further investigations. Nonetheless, nonlinear bias is more effectively constrained by 3PCF modelling, which will be discussed in the subsequent Section.

b_1	b_2	b_t
$\mathcal{U}(0,5)$	$\mathcal{U}(-5,10)$	$\mathcal{U}(-2,5)$

Table 6.7: *Prior distributions for the parameters used in the ELM 3PCF analysis, where* $\mathcal{U}(a,b)$ *indicates a uniform distribution between* a *and* b*.*

6.3.2 Three-point correlation function modelling

We then performed the 3PCF modelling, fitting the mean signal in the fixed sides configuration, with $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc, from the ELM using the B23 model, described in Section 3.4. The posterior distribution was sampled using a MCMC with 64 walkers, with a length of 8000. The model parameters included b_1, b_2, and b_t, where b_1 denotes the linear bias, b_2 represents the nonlinear bias, and b_t is the tidal bias. As previously discussed, the modelling of the BS does not depend on the amplitude of the matter density fluctuations, σ_8. All parameters were allowed to vary freely with uniform prior distributions, as shown in Table 6.7.

Figure 6.10 displays the data alongside the best-fit model for both the total and pure samples across the two redshift ranges. The fitting range included all values of the angular variable, θ, in the interval $[0, \pi]$. The model accurately reproduced the connected 3PCF, with discrepancies observed only at low values of θ. These discrepancies correspond to small scales for the third side of the triangular configuration, which are influenced by nonlinear behavior.

Figures 6.11 and 6.12 present the 1D and 2D posterior distributions for the total and pure fits in the first and second redshift ranges, respectively. The best-fit values and 68% confidence regions are summarized in Tables 6.8 and 6.9. In this case, the linear bias best-fit value, b_1, exhibited attenuation due to interloper contamination in both redshift ranges, specifically by 20.3% and 24.3%, respectively. However, the posterior distributions for the nonlinear bias, b_2, and tidal bias, b_t, did not appear to be strongly affected by systematic effects.

These results align with those obtained from the 2PCF modelling, particularly regarding the best-fit values of b_1 and b_2. However, we did not observe the same significant difference in the 68% confidence regions of the nonlinear bias between the total and pure posteriors. This suggests that the observed feature may be specific to the PS modeling or may not be visible at these particular scales for the BS.

6.3.3 Towards probes combination

The next step to achieve tighter constraints on the biases of the cosmological parameters due to interloper contamination is to carry out a joint analysis of the 2PCF and 3PCF. By integrating information from both the 2PCF and 3PCF measurements, we

Model	b_1	b_2	b_t
Total sample	$1.37^{+0.18}_{-0.20}$	$2.1^{+1.9}_{-1.3}$	$1.03^{+0.96}_{-0.65}$
Pure sample	$1.72^{+0.18}_{-0.20}$	$2.0^{+1.7}_{-1.2}$	$0.98^{+0.87}_{-0.62}$

Table 6.8: *Best-fit values and 68% confidence region for the free parameters b_1, b_2 and b_t of the B23 model, obtained from the analysis of the total (first row) and pure (second row) ELM connected 3PCF in the redshift range* $0.9 < z < 1.1$.

Model	b_1	b_2	b_t
Total sample	$1.34^{+0.19}_{-0.21}$	$2.3^{+2.1}_{-1.4}$	$0.86^{+0.98}_{-0.64}$
Pure sample	$1.77^{+0.21}_{-0.23}$	$2.6^{+2.1}_{-1.5}$	$0.96^{+0.98}_{-0.70}$

Table 6.9: *Best-fit values and 68% confidence region for the free parameters b_1, b_2 and b_t of the B23 model, obtained from the analysis of the total (first row) and pure (second row) ELM connected 3PCF in the redshift range* $1.1 < z < 1.3$.

can estimate the model parameters more accurately. For example, this approach exploits the PS model's ability to constrain the linear bias and the BS model's strength in tightening constraints on the nonlinear bias. Figures 6.13, 6.14, 6.15, and 6.16 illustrate how constraints on b_1 and b_2 should improve. These figures compare the b_1 and b_2 posterior distributions obtained from the 2PCF and 3PCF modeling. For the 2PCF, we normalized both the linear and nonlinear bias posterior distributions by σ_8 to facilitate comparison with those from the 3PCF statistics. The values for the amplitude of the matter density fluctuations, σ_8, computed using a CBL function with *CAMB* (Lewis et al., 2000 [50]), are $\sigma_8 = 0.502$ for the first redshift range and $\sigma_8 = 0.461$ for the second redshift range.

The four plots demonstrate that the posterior distributions are compatible, and their intersection should be approximately the constraints on b_1 and b_2 from joint modelling. The first two plots, which show the results of the pure samples modelling, confirm that the 2PCF places tighter constraints on b_1, whereas the 3PCF more effectively constrains b_2. The last two plots, representing the total sample modelling, confirm the tighter constraints on b_1 by 2PCF, but indicate comparable constraints on b_2 by both 2PCF and 3PCF. This comparable constraining power for b_2 in the total samples is due to the shrinking of the confidence region of the nonlinear bias parameter for samples contaminated by redshift interlopers, as discussed in Paragraph 6.3.1. This intriguing feature warrants further investigations.

In summary, the results reveal a substantial offset between the overlap of the 2D confidence regions of the total and pure results, primarily due to the bias in the b_1 parameter.

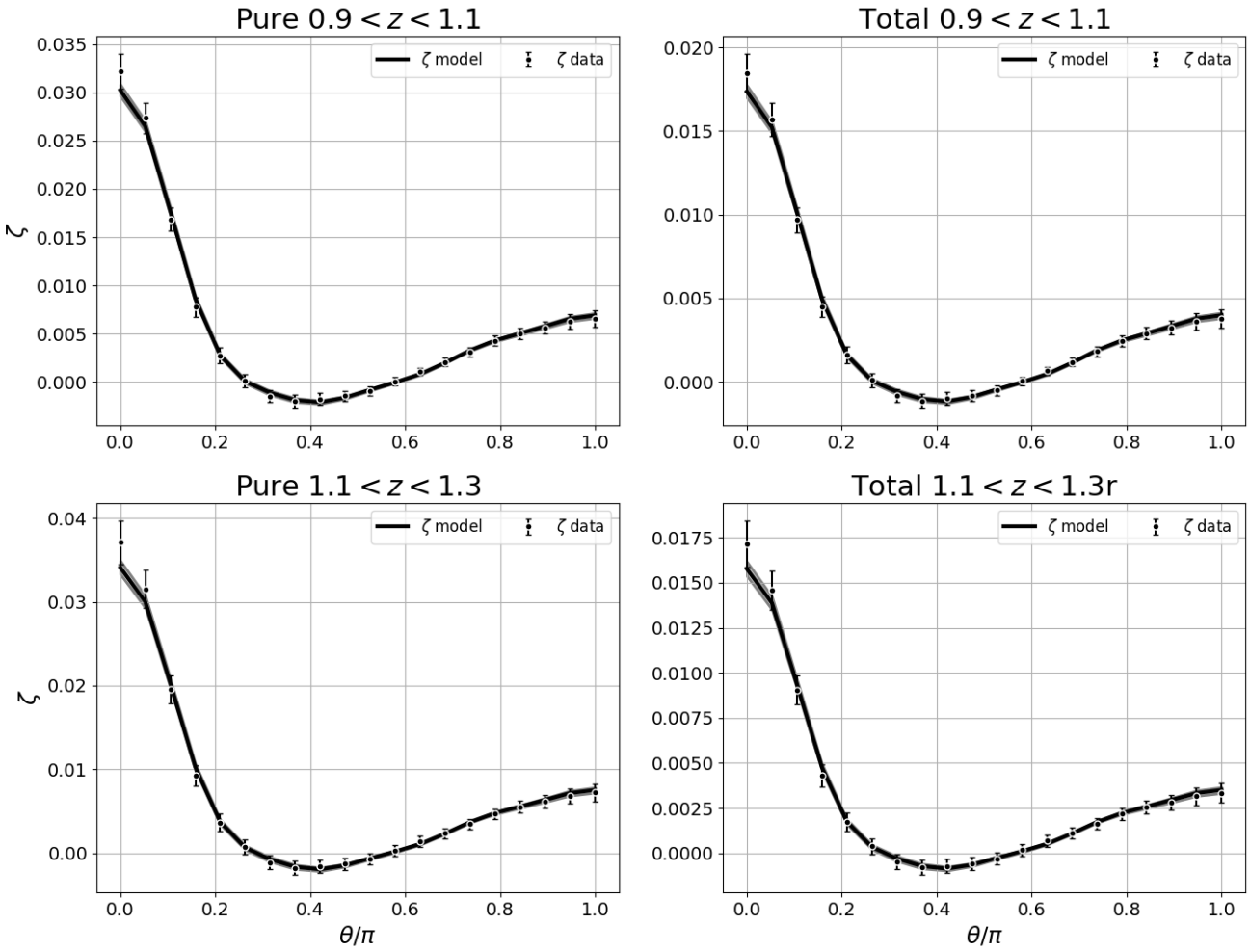


Figure 6.10: *Total (right panels) and pure (left panels) ELM 3PCF signal (in black) in the first (upper panels) and second (lower panels) redshift range. Circles represent the data. The error bars are the square root of the diagonal elements of the covariance matrix. The solid curves show the best-fit model, while the grey shaded areas represent the 68% confidence region of the model.*

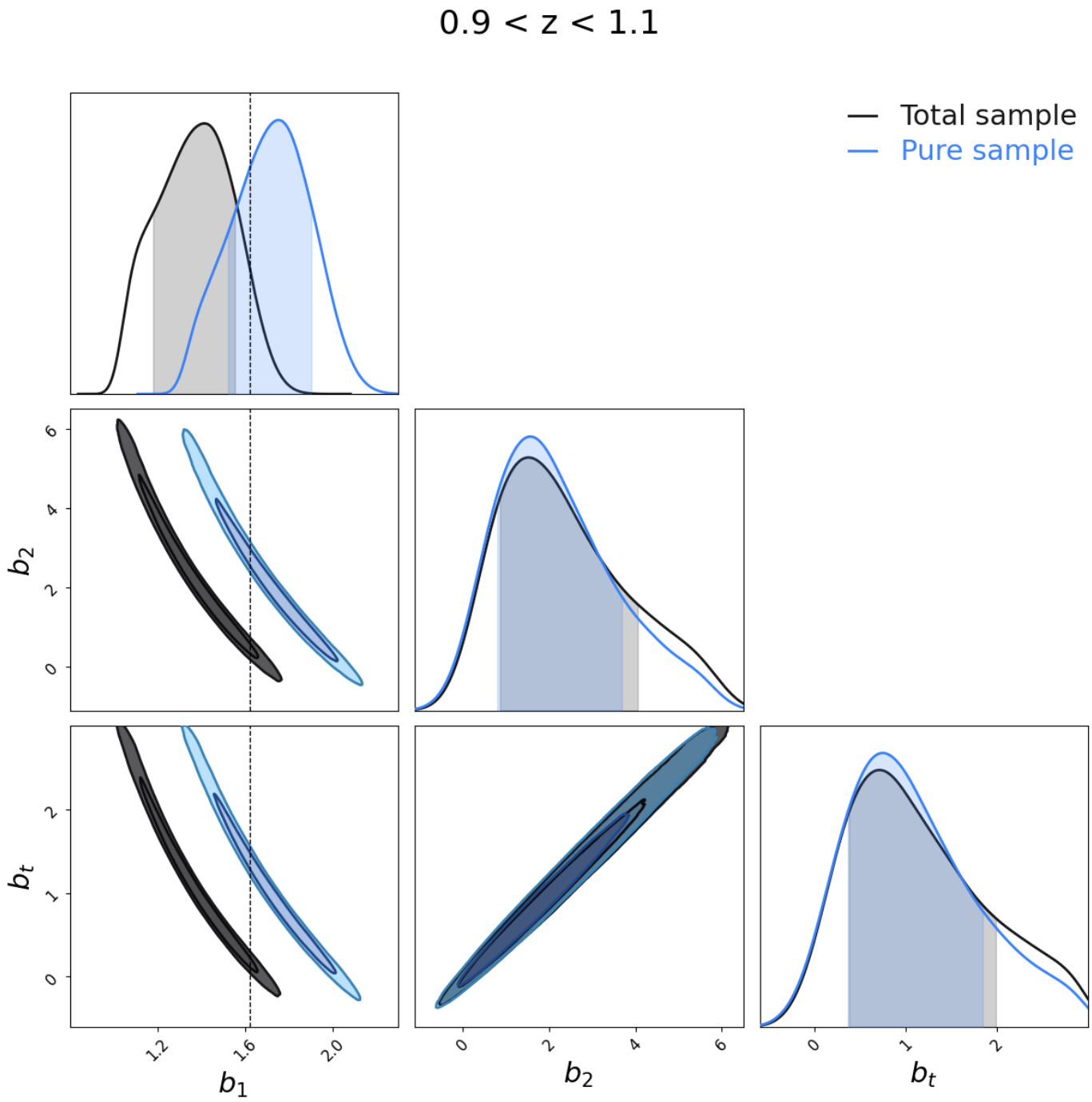


Figure 6.11: *1D and 2D posterior distributions for the free parameters b_1, b_2 and b_t of the B23 model, obtained from the analysis of the total (in orange) and pure (in blue) ELM connected 3PCF in the redshift range $0.9 < z < 1.1$. The darker and lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

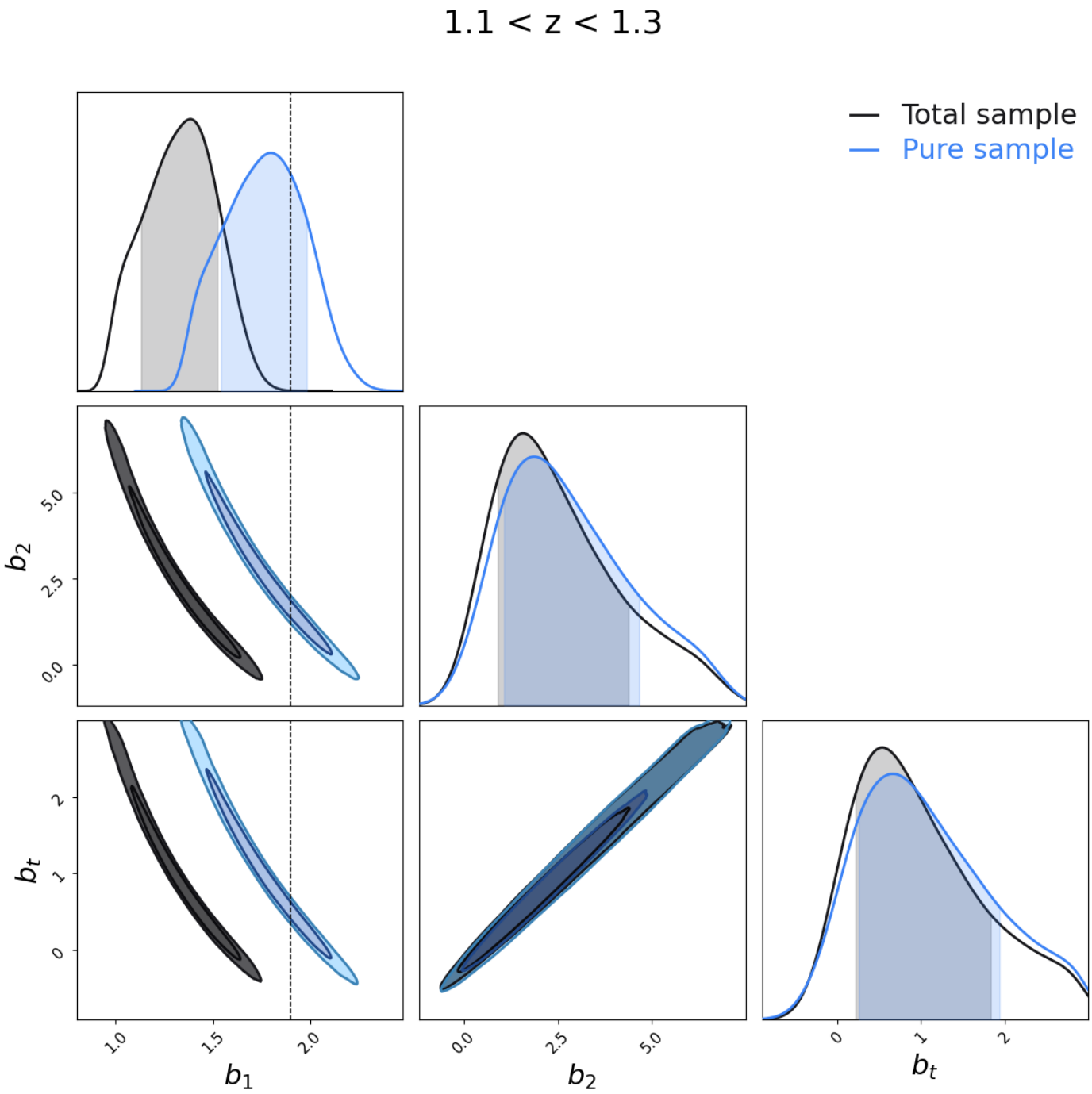


Figure 6.12: *1D and 2D posterior distributions for the free parameters* b_1, b_2 *and* b_t *of the B23 model, obtained from the analysis of the total (in orange) and pure (in blue) ELM connected 3PCF in the redshift range* $1.1 < z < 1.3$. *The darker and lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

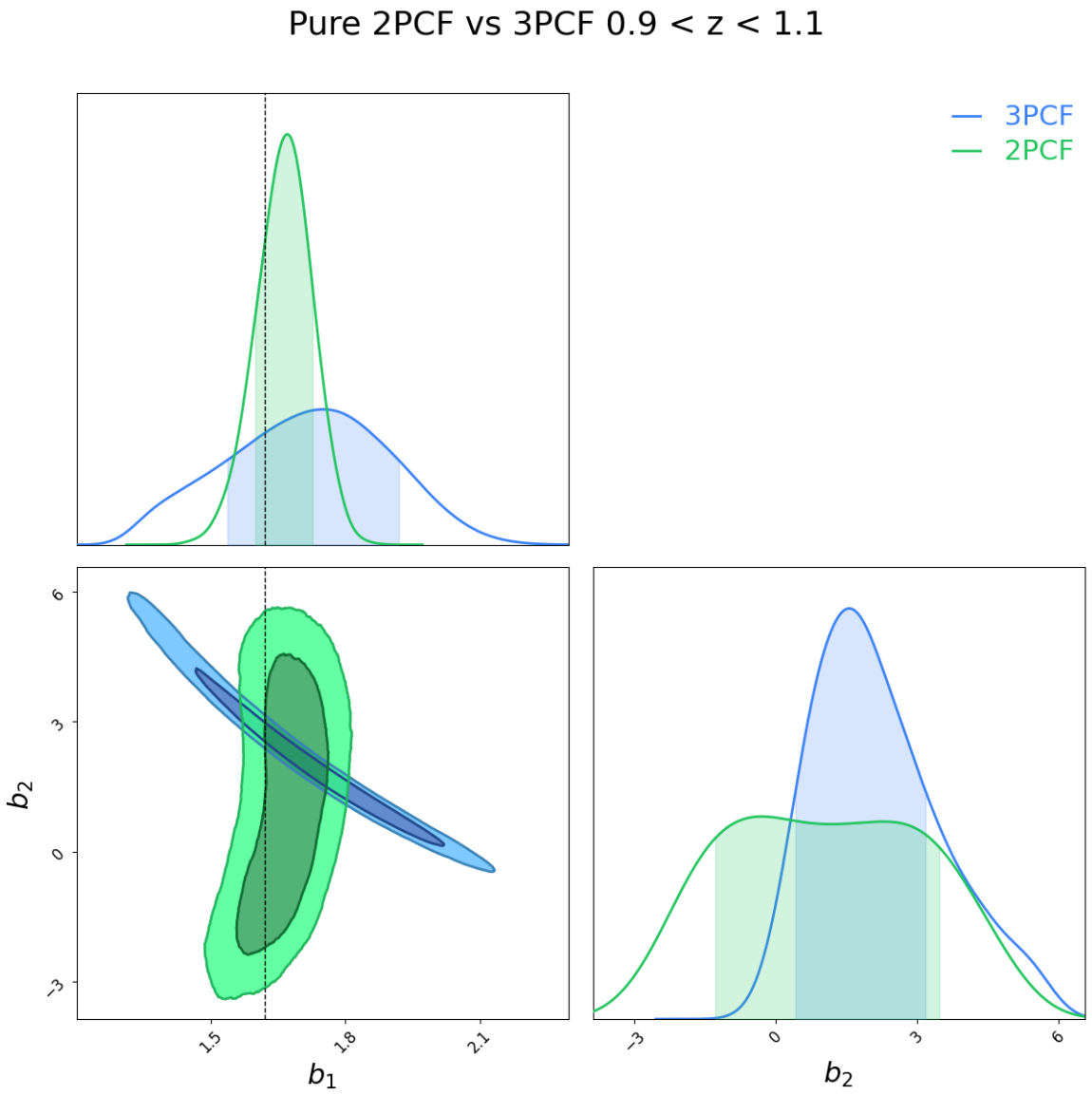


Figure 6.13: *1D and 2D posterior distributions for the parameters b1 and b2 obtained from the analyses of the pure 2PCF (green) and 3PCF (blue) in the redshift range* $0.9 < z < 1.1$. *The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

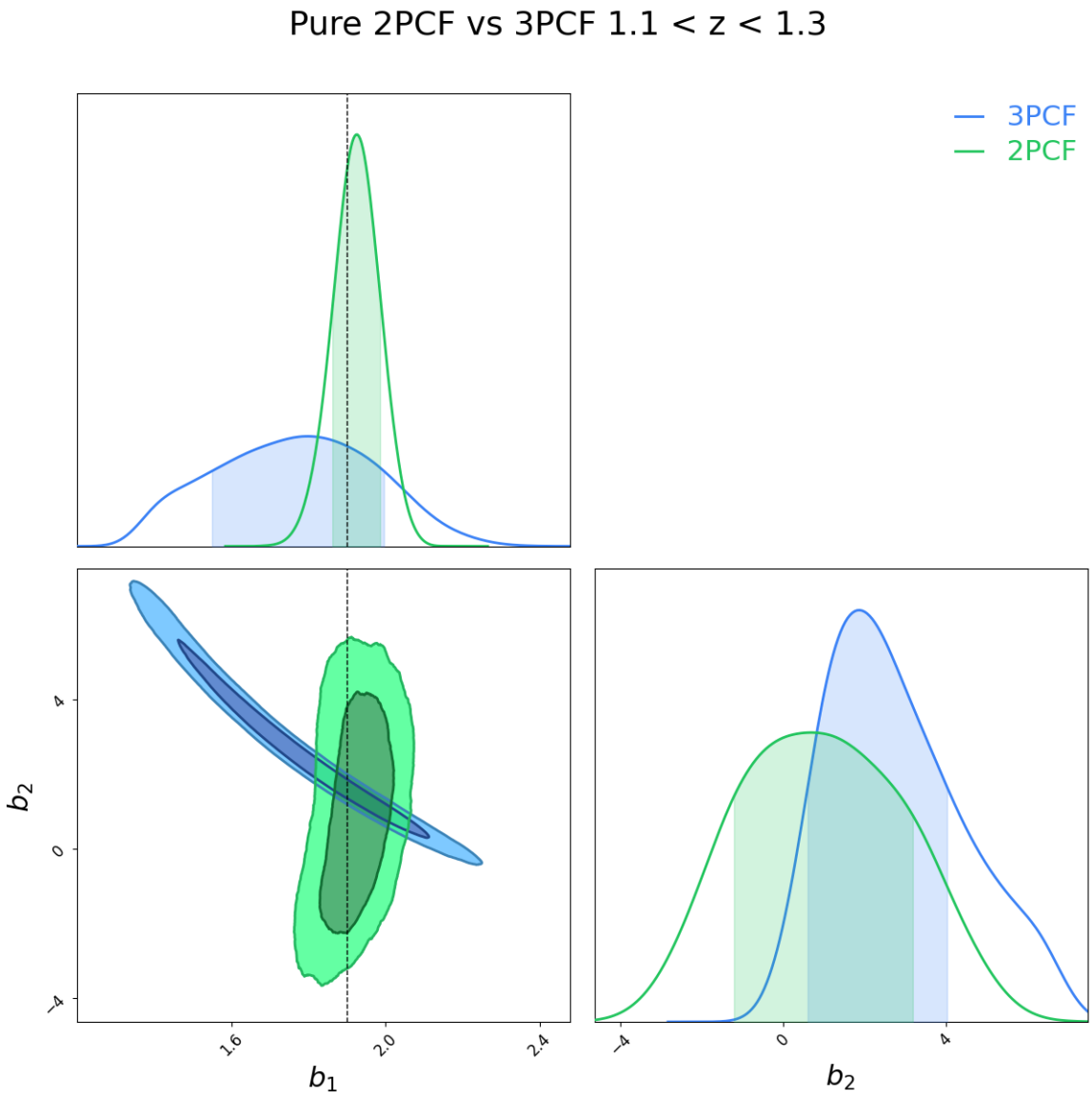


Figure 6.14: *1D and 2D posterior distributions for the parameters b1 and b2 obtained from the analyses of the pure 2PCF (green) and 3PCF (blue) in the redshift range* $1.1 < z < 1.3$. *The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

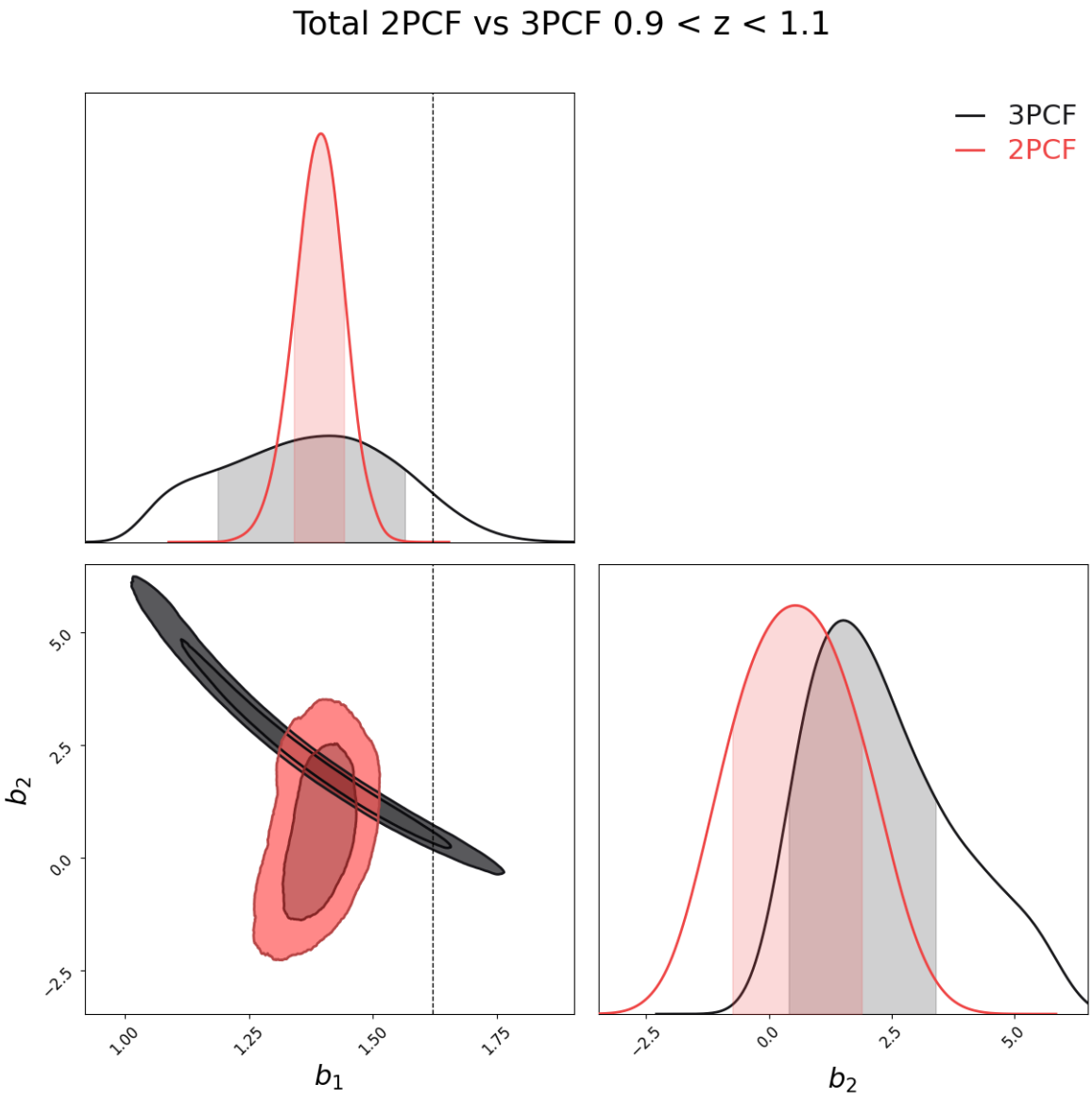


Figure 6.15: *1D and 2D posterior distributions for the parameters b1 and b2 obtained from the analyses of the total 2PCF (red) and 3PCF (black) in the redshift range* $0.9 < z < 1.1$. *The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

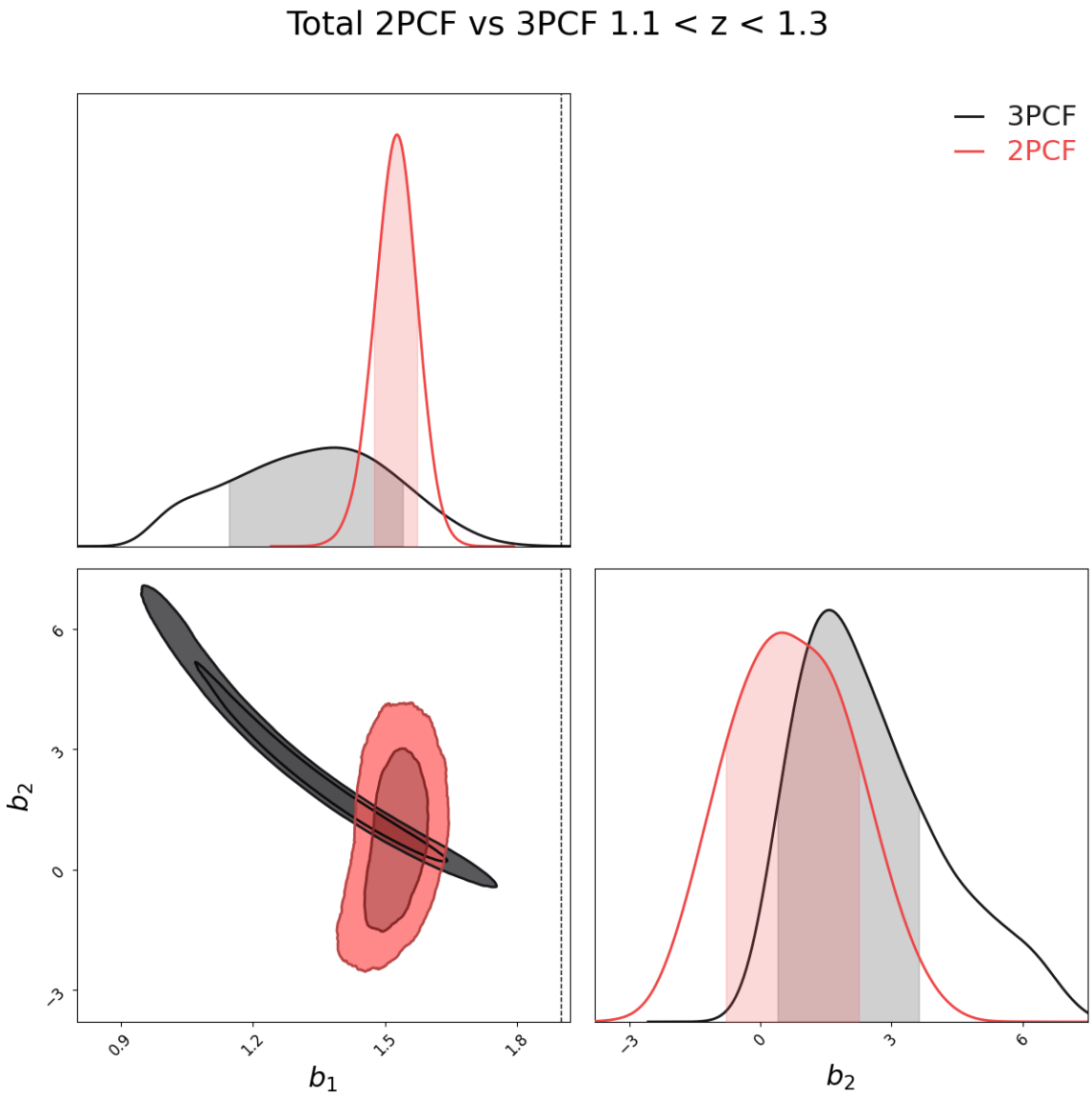


Figure 6.16: *1D and 2D posterior distributions for the parameters b1 and b2 obtained from the analyses of the total 2PCF (red) and 3PCF (black) in the redshift range* $1.1 < z < 1.3$. *The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed black lines represent the theoretical predictions for the parameters.*

Chapter 7

Conclusions

During the last decades, exponential progress has been registered in the study of LSS, which brought to the establishment of the ΛCDM model. Despite this, significant questions still remain open, particularly regarding the inflationary scenario, the nature of cold dark matter, and the cause of the Universe's accelerated expansion. Clustering and the study of LSS are pivotal probes for constraining cosmology, and they will play a central role in upcoming missions such as ESA's *Euclid*.

The *Euclid* wide survey will cover over $15000\,\mathrm{deg}^2$ of the sky, measuring galaxy positions and estimating their spectroscopic redshifts, primarily using the $H\alpha$ line. This mission aims to place unprecedented constraints on the dark energy equation of state parameter, w_Λ, test the validity of GR on cosmological scales (constraining the growth index parameter γ as discussed in Chapter 1), and constrain the mass of neutrinos and non-Gaussian initial conditions of the matter density field [7]. However, the low SNR of redshift measurements, stemmed as a compromise to ensure a deep enaugh survey, introduces systematic errors which will affect redshift measurements, such as redshift interlopers. These systematic uncertainties will subsequently impact the results derived from statistical analysis performed to constrain cosmological parameters, through e.g. n-point correlation functions and polyspectra.

This study aimed to quantify the impact of redshift interlopers on the measurement and modeling of the 2PCF and the 3PCF and to initiate the development of mitigation strategies. Such studies are essential to fully exploit the capabilities of *Euclid*. This Chapter summarizes the main findings of our work, and discusses future perspectives.

7.1 Main results

The most important results of this study are the following:

1. Firstly, the derivation of a new estimator for the total 3PCF signal of a catalogue, which can separate the contributions of the self- and cross-correlations of its n

mutually exclusive sub-catalogues. The expressions for these correlations are given as follows (as discussed in Section 3.3.1):

$$\hat{\zeta}_{tot} = \sum_i f_i^3 \frac{R_i R_i R_i}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{iii} + 3 \left[\sum_{i \neq j, i<j} \left(f_i^2 f_j \frac{R_i R_i R_j}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{iij} + f_i f_j^2 \frac{R_i R_j R_j}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{ijj} \right) \right] + 6 \sum_{i \neq j \neq k, i<j<k} f_i f_j f_k \frac{R_i R_j R_k}{R_{tot} R_{tot} R_{tot}} \hat{\zeta}_{ijk}, \tag{7.1}$$

$$\hat{\zeta}_{iij} = \frac{D_i D_i D_j - D_i D_i R_j - 2 D_i R_i D_j + 2 D_i R_i R_j + R_i R_i D_j - R_i R_i R_j}{R_i R_i R_j}, \tag{7.2}$$

$$\hat{\zeta}_{ijj} = \frac{D_i D_j D_j + D_i R_j R_j - 2 D_i D_j R_j + 2 R_i D_j R_j - R_i D_j D_j - R_i R_j R_j}{R_i R_j R_j}, \tag{7.3}$$

$$\hat{\zeta}_{ijk} = \frac{D_i D_j D_k - D_i R_j D_k - D_i D_j R_k - R_i D_j D_k + R_i R_j D_k + R_i D_j R_k + D_i R_j R_k - R_i R_j R_k}{R_i R_j R_k} \tag{7.4}$$

where $\hat{\zeta}_{tot}$ is the total 3PCF estimator, $\hat{\zeta}_{iii}$, $\hat{\zeta}_{iij}$ and $\hat{\zeta}_{ijk}$ are the self- bi-cross- and tri-cross- estimators for the 3PCF, respectively (described in Paragraph 3.3.1). Moreover, f_i is the ratio of the number of tracers in the i-th sub-catalogue to the total number of tracers (Eq. 3.16) and DDD, DDR, DRR, and RRR are actually $DDD(r_{12}, r_{13}, \theta)$, $DDR(r_{12}, r_{13}, \theta)$, $DRR(r_{12}, r_{13}, \theta)$, and $RRR(r_{12}, r_{13}, \theta)$, which represent the number of normalized counts of data-data-data, data-data-random, data-random-random, and random-random-random triplets, respectively, in the triangular configuration (r_{12}, r_{13}, θ). The pedex $_i$ indicates from which catalogue the tracer comes from. A more detailed description can be found in Paragraph 3.3.1.

2. Subsequently, we implemented a total of four new classes and nineteen new functions in the CBL to measure:

   - The monopole of the cross-correlations of the 2PCF with the LS estimator using two random catalogues and its Poissonian error (Section 4.1).
   - The first three even multipoles of the cross-correlations of the 2PCF with the LS integrated estimator using two random catalogues and their Poissonian error (Section 4.2).

- The self- and cross-correlations of the connected 3PCF using our newly proposed estimators and their Poissonian error (Section 4.3).

These functions were thoroughly tested, with details provided in Chapter 4.

3. Upon validating these statistical tools, we employed them to assess the impact of redshift interlopers on the 2PCF and 3PCF measurements from the FS2 catalogue. The results, summarized in Tables 5.2 and 5.3, indicate that the pure 2PCF monopole signal is attenuated by 25-40% at small and intermediate scales ($20 - 120\,h^{-1}$Mpc), with higher attenuation at larger scales ($150 - 180\,h^{-1}$Mpc), reaching a peak of 78% for the redshift range $0.9 < z < 1.1$. At higher redshifts ($1.5 < z < 1.8$), the impact of interlopers is lower, with 27.8% and 25.5% damping at small and large scales, respectively. For the first time, we extended the study to three-point statistics. As reported in Table 5.3, the damping of the 3PCF signal is relatively constant for different values of the angular variable θ, assuming values mostly in the range of 40-60%. Additionally, we notice that the impact is lower at higher redshifts.

4. Furthermore, a residual analysis validated the expression for the total 2PCF and the newly derived expression for the total 3PCF signal in terms of the self- and cross-correlations of n mutually exclusive sub-catalogues (Figures 5.8 and 5.13). The importance of cross-correlations in retrieving the total 2PCF signal was highlighted, showing non-negligible contributions except at higher redshifts ($z > 1.5$). Overall, we demonstrated that at high redshifts, the impact of interlopers diminishes, both in damping of the signal and in the significance of cross-correlations.

5. Finally, for the first time we extended the study to the cosmological analysis. We showed that the interloper contamination significantly biases the constraints on the main cosmological parameters. From the 2PCF modeling, we found that the growth factor $f\,\sigma_8$, estimated from the contaminated catalogue signals, is attenuated by 25.8% and 28.7% in the $0.9 < z < 1.1$ and $1.1 < z < 1.3$ ranges, respectively, compared to the pure galaxy signal, which matches the theoretical predictions. Additionally, the linear bias parameter $b_1\,\sigma_8$ suffers biases of 16.3% and 20.7% due to interloper contamination. These findings were corroborated by the 3PCF modeling, which showed an attenuation of 20.1% and 24.6% for the linear bias parameter.

In summary, this study has demonstrated that interloper contamination significantly affects both the measurement and modeling of the 2PCF and 3PCF, especially at lower redshifts, with impacts reaching up to 70% of the measured signal and 30% of the estimated cosmological parameters. Therefore, it is crucial to continue the research in this area to develop effective mitigation strategies.

7.2 Future perspectives

The natural continuation of this study consists of extending the analysis of the impact of redshift interlopers to the modelling of the 3PCF at all scales (thus using simultaneously more r_{12} and r_{13} combinations to set constraints on the cosmological parameters) and the joint modelling of 2PCF and 3PCF, as discussed in Paragraph 6.3.3. These approaches will allow us to place tighter constraints on parameter biases and represent the next step of our work.

Furthermore, some intriguing features highlighted by our analysis require further investigations. For example, the peculiar behavior of the $b_2\,\sigma_8$ parameter in the modeling of the 2PCF, which appears to improve the posterior distribution for the contaminated sample, needs to be addressed in future studies.

Ultimately, the goal of this research is to incorporate the modeling of interlopers into the likelihood to fully recover the true signal from the total contaminated 2PCF and 3PCF. This is crucial because we have demonstrated that redshift interlopers significantly impact the posterior distributions of cosmological parameters in both 2PCF and 3PCF modelling. In this context, our work is fundamental as we developed new tools to study the 3PCF signal and constrained the importance of cross-correlations in recovering the total signal.

Acknowledgements

After all this talk about galaxies, allow me to thank some people with a few clichés, in what is the only part of my thesis that will be read by most.

First of all, my supervisors, Prof. Michele Moresco, Prof. Federico Marulli, and Prof. Sylvain de la Torre, for all they have taught me, as well as for their kindness and availability in times of need. A big thank you also goes to the University of Bologna (UNIBO) for helping me grow as an astrophysicist and to the Laboratory of Astrophysics of Marseille (LAM) for hosting me during the first months of my thesis project.

My family. My mother Teresa, for her patience and her love. My father Andrea, to whom I owe my passion for the Universe. My brother Ernesto, who helped me taking my mind off work. Diana and Lucy, for their affection. Nonno Giuliano, with whom I shared carbonare and GPs between exams. Stella, all the uncles and aunts, and the grandmothers, for always supporting and hosting me, making me feel at home.

Eugenio. Calling you a 'friend' only scratches the surface of the concept. Wherever we end up, we will never be alone.

Laura. Your presence in Bologna these years has been as precious as my mental health.

Bologna. Alessandro, Ilaria, Leonardo, Lorenzo, Lorenzo, Niccolò, Nicolò (yes, I put you in alphabetical order) and the other course mates. I could not have chosen better company to share the journey, the city, and the beers with.

Marseille. To the city, you are the one of my dreams. To Meriam, Duarte, Martin and everyone else, thank you for welcoming and helping me.

Alimathà. The two months I spent on the island were the best of my life. The people I met there will forever be in my heart. An enchanting bubble.

The violin, AC Milan, beer.

Ringraziamenti

Dopo tutto questo parlare di galassie, permettetemi di ringraziare qualche persona con un po' di frasi fatte, in quella che è l'unica parte della mia tesi che verrà letta dai più.

Prima di tutto i miei relatori, Prof. Michele Moresco, Prof. Federico Marulli, e Prof. Sylvain de la Torre, per tutto ciò che mi hanno insegnato, oltre che per la loro gentilezza e disponibilità nel momento del bisogno. Un gran ringraziamento va anche all'Università di Bologna (UNIBO) per avermi aiutato a crescere come astrofisico e al Laboratorio di Astrofisica di Marsiglia (LAM) per avermi ospitato nei primi mesi del mio progetto di tesi.

La mia famiglia. Mia madre Teresa, per la sua pazienza e il suo amore. Mio padre Andrea, a cui devo la mia passione per l'Universo. Mio fratello Ernesto, per avermi aiutato a svagarmi con momenti di leggerezza. Diana e Lucy, per il loro affetto. Nonno Giuliano, con cui ho condiviso carbonare e GP tra un esame e l'altro. Stella, tutti gli zii e le nonne, per avermi sempre supportato e ospitato, facendomi sentire a casa.

Eugenio. Definirti 'amico' sfiora soltanto il concetto. Ovunque finiremo, non saremo mai soli.

Laura. La tua presenza a Bologna in questi anni è stata preziosa quanto la mia sanità mentale.

Bologna. Alessandro, Ilaria, Leonardo, Lorenzo, Lorenzo, Niccolò, Nicolò (sì, vi ho messi in ordine alfabetico) e gli altri compagni di corso. Non avrei potuto scegliere compagnia migliore per condividere il percorso, la città e le birre.

Marsiglia. Alla città, sei quella dei miei sogni. A Meriam, Duarte, Martin e tutti gli altri, grazie per avermi accolto e aiutato.

Alimathà. I due mesi che ho passato sull'isola sono stati i più belli della mia vita. Le persone che vi ho conosciuto saranno per sempre nel mio cuore. Una bolla incantevole.

Il violino, l'AC Milan, la birra.

Bibliography

[1] Federico Marulli, Alfonso Veropalumbo, and Michele Moresco. "CosmoBolognaLib: C++ libraries for cosmological calculations". In: *Astronomy and Computing* 14 (2015), pp. 35–42. eprint: Arxiv:1511.00012v2. URL: https://arxiv.org/abs/1511.00012.

[2] Yuting Wang, Lixin Xu, and Gong-Bo Zhao. "A Measurement of the Hubble Constant Using Galaxy Redshift Surveys". In: *The Astrophysical Journal* 849.2 (Nov. 2017), p. 84. DOI: 10.3847/1538-4357/aa8f48. URL: https://dx.doi.org/10.3847/1538-4357/aa8f48.

[3] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: 641, A6 (Sept. 2020), A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[4] Lucy Calder and Ofer Lahav. "Dark energy: how the paradigm shifted". In: *Physics World* 23.01 (Jan. 2010), p. 32. DOI: 10.1088/2058-7058/23/01/33. URL: https://dx.doi.org/10.1088/2058-7058/23/01/33.

[5] Eugenio Bianchi and Carlo Rovelli. *Why all these prejudices against a constant?* 2010. arXiv: 1002.3966 [astro-ph.CO].

[6] Albert Einstein. "Cosmological Considerations in the General Theory of Relativity." In: *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math.Phys.)* (1917), pp. 142–152.

[7] R. Laureijs et al. *Euclid Definition Study Report*. 2011. arXiv: 1110.3193 [astro-ph.CO].

[8] Jeremy L. Tinker et al. "THE LARGE-SCALE BIAS OF DARK MATTER HALOS: NUMERICAL CALIBRATION AND MODEL TESTS". In: *The Astrophysical Journal* 724.2 (Nov. 2010), pp. 878–886. ISSN: 1538-4357. DOI: 10.1088/0004-637x/724/2/878. URL: http://dx.doi.org/10.1088/0004-637X/724/2/878.

[9] Neal Katz, Lars Hernquist, and David H. Weinberg. "The Clustering of High-Redshift Galaxies in the Cold Dark Matter Scenario". In: *The Astrophysical Journal* 523.2 (Oct. 1999), p. 463. DOI: 10.1086/307744. URL: https://dx.doi.org/10.1086/307744.

[10] Renyue Cen and Jeremiah P. Ostriker. "Physical Bias of Galaxies from Large-Scale Hydrodynamic Simulations". In: *The Astrophysical Journal* 538.1 (July 2000), p. 83. DOI: 10.1086/309090. URL: https://dx.doi.org/10.1086/309090.

[11] J. N. Fry and Enrique Gaztanaga. "Biasing and Hierarchical Statistics in Large-Scale Structure". In: 413 (Aug. 1993), p. 447. DOI: 10.1086/173015. arXiv: astro-ph/9302009 [astro-ph].

[12] Masahiro Takada and Bhuvnesh Jain. "The three-point correlation function in cosmology". In: *Monthly Notices of the Royal Astronomical Society* 340.2 (Apr. 2003), pp. 580–608. ISSN: 1365-2966. DOI: 10.1046/j.1365-8711.2003.06321.x. URL: http://dx.doi.org/10.1046/j.1365-8711.2003.06321.x.

[13] István Szapudi. *Introduction to Higher Order Spatial Statistics in Cosmology*. 2005. arXiv: astro-ph/0505391 [astro-ph].

[14] F. Bernardeau et al. "Large-scale structure of the Universe and cosmological perturbation theory". In: *Physics Reports* 367.1–3 (Sept. 2002), pp. 1–248. ISSN: 0370-1573. DOI: 10.1016/s0370-1573(02)00135-7. URL: http://dx.doi.org/10.1016/S0370-1573(02)00135-7.

[15] P. J. E. Peebles and E. J. Groth. "Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog." In: 196 (Feb. 1975), pp. 1–11. DOI: 10.1086/153390.

[16] E. J. Groth and P. J. E. Peebles. "Statistical analysis of catalogs of extragalactic objects. VII. Two- and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies." In: 217 (Oct. 1977), pp. 385–405. DOI: 10.1086/155588.

[17] Michele Moresco et al. "C3: Cluster Clustering Cosmology. ii. First Detection of the Baryon Acoustic Oscillations Peak in the Three-point Correlation Function of Galaxy Clusters". In: *The Astrophysical Journal* 919.2 (Oct. 2021), p. 144. DOI: 10.3847/1538-4357/ac10c9. URL: https://dx.doi.org/10.3847/1538-4357/ac10c9.

[18] Y. P. Jing, G. Börner, and R. Valdarnini. "Three-point correlation function of galaxy clusters in cosmological models – a strong dependence on triangle shapes". In: *Monthly Notices of the Royal Astronomical Society* 277.2 (Nov. 1995), pp. 630–640. ISSN: 0035-8711. DOI: 10.1093/mnras/277.2.630. eprint: https://academic.oup.com/mnras/article-pdf/277/2/630/18200179/mnras277-0630.pdf. URL: https://doi.org/10.1093/mnras/277.2.630.

[19] E. Gaztañaga and R. Scoccimarro. "The three-point function in large-scale structure: redshift distortions and galaxy bias". In: *Monthly Notices of the Royal Astronomical Society* 361.3 (Aug. 2005), pp. 824–836. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09234.x. eprint: https://academic.oup.com/mnras/article-pdf/361/3/824/2940536/361-3-824.pdf. URL: https://doi.org/10.1111/j.1365-2966.2005.09234.x.

[20] Felipe Marín. "THE LARGE-SCALE THREE-POINT CORRELATION FUNCTION OF SLOAN DIGITAL SKY SURVEY LUMINOUS RED GALAXIES". In: *The Astrophysical Journal* 737.2 (Aug. 2011), p. 97. DOI: 10.1088/0004-637X/737/2/97. URL: https://dx.doi.org/10.1088/0004-637X/737/2/97.

[21] Yu Wang et al. "The three-point correlation function of galaxies: comparing halo occupation models with observations". In: *Monthly Notices of the Royal Astronomical Society* 353.1 (Sept. 2004), pp. 287–300. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2004.08141.x. eprint: https://academic.oup.com/mnras/article-pdf/353/1/287/18658464/353-1-287.pdf. URL: https://doi.org/10.1111/j.1365-2966.2004.08141.x.

[22] Enrique Gaztañaga et al. "Clustering of luminous red galaxies – III. Baryon acoustic peak in the three-point correlation". In: *Monthly Notices of the Royal Astronomical Society* 399.2 (Oct. 2009), pp. 801–811. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.15313.x. eprint: https://academic.oup.com/mnras/article-pdf/399/2/801/3639272/mnras0399-0801.pdf. URL: https://doi.org/10.1111/j.1365-2966.2009.15313.x.

[23] Scott Dodelson and Fabian Schmidt. "Modern Cosmology". In: (Nov. 2003).

[24] Daniel J. Eisenstein et al. "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *The Astrophysical Journal* 633.2 (Nov. 2005), pp. 560–574. ISSN: 1538-4357. DOI: 10.1086/466512. URL: http://dx.doi.org/10.1086/466512.

[25] Elena Massara et al. "Line confusion in spectroscopic surveys and its possible effects: shifts in Baryon Acoustic Oscillations position". In: *Monthly Notices of the Royal Astronomical Society* 508.3 (Sept. 2021), pp. 4193–4201. ISSN: 0035-8711. DOI: 10.1093/mnras/stab2628. eprint: https://academic.oup.com/mnras/article-pdf/508/3/4193/40825154/stab2628.pdf. URL: https://doi.org/10.1093/mnras/stab2628.

[26] Anthony R. Pullen et al. "Interloper bias in future large-scale structure surveys". In: *Publications of the Astronomical Society of Japan* 68.1 (Dec. 2015), p. 12. ISSN: 0004-6264. DOI: 10.1093/pasj/psv118. eprint: https://academic.oup.com/pasj/article-pdf/68/1/12/54679927/pasj_68_1_12.pdf. URL: https://doi.org/10.1093/pasj/psv118.

[27] G. E. Addison et al. “The Impact of Line Misidentification on Cosmological Constraints from Euclid and Other Spectroscopic Galaxy Surveys”. In: 879.1, 15 (July 2019), p. 15. DOI: 10.3847/1538-4357/ab22a0. arXiv: 1811.10668 [astro-ph.CO].

[28] Henry S. Grasshorn Gebhardt et al. “Unbiased Cosmological Parameter Estimation from Emission-line Surveys with Interlopers”. In: 876.1, 32 (May 2019), p. 32. DOI: 10.3847/1538-4357/ab12d5. arXiv: 1811.06982 [astro-ph.CO].

[29] P. J. E. Peebles and M. G. Hauser. “Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs”. In: 28 (Nov. 1974), p. 19. DOI: 10.1086/190308.

[30] M. Davis and P. J. E. Peebles. “A survey of galaxy redshifts. V. The two-point position and velocity correlations.” In: 267 (Apr. 1983), pp. 465–482. DOI: 10.1086/160884.

[31] A. J. S. Hamilton. “Toward Better Ways to Measure the Galaxy Correlation Function”. In: 417 (Nov. 1993), p. 19. DOI: 10.1086/173288.

[32] Stephen D. Landy and Alexander S. Szalay. “Bias and Variance of Angular Correlation Functions”. In: 412 (July 1993), p. 64. DOI: 10.1086/172900.

[33] Antoine Labatie et al. “Uncertainty in 2-point correlation function estimators and baryon acoustic oscillation detection in galaxy surveys”. In: *Statistical Methodology* 9.1 (2012). Special Issue on Astrostatistics + Special Issue on Spatial Statistics, pp. 85–100. ISSN: 1572-3127. DOI: https://doi.org/10.1016/j.stamet.2011.05.001. URL: https://www.sciencedirect.com/science/article/pii/S1572312711000499.

[34] István Szapudi and Alexander S. Szalay. “A New Class of Estimators for the N-point Correlations”. In: (1997). eprint: Arxiv:astro-ph/9704241v1. URL: https://arxiv.org/abs/astro-ph/9704241.

[35] Paolo Catelan et al. “The bias field of dark matter haloes”. In: *Monthly Notices of the Royal Astronomical Society* 297.3 (July 1998), pp. 692–712. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1998.01455.x. eprint: https://academic.oup.com/mnras/article-pdf/297/3/692/3024155/297-3-692.pdf. URL: https://doi.org/10.1046/j.1365-8711.1998.01455.x.

[36] Florian Beutler et al. “The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles”. In: *Monthly Notices of the Royal Astronomical Society* 443 (Sept. 2014), p. 1065. eprint: Arxiv:1312.4611v2. URL: https://arxiv.org/abs/1312.4611.

[37] Atsushi Taruya, Takahiro Nishimichi, and Shun Saito. "Baryon Acoustic Oscillations in 2D: Modeling Redshift-space Power Spectrum from Perturbation Theory". In: *Physical Review D* 82 (2010), p. 063522. eprint: Arxiv:1006.0699v1. URL: https://arxiv.org/abs/1006.0699.

[38] Nick Kaiser. "Clustering in real space and in redshift space". In: *Monthly Notices of the Royal Astronomical Society* 227.1 (1987), pp. 1–21. URL: https://doi.org/10.1093/mnras/227.1.1.

[39] Shun Saito et al. "Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum". In: *Physical Review D* 90.12 (Dec. 2014). ISSN: 1550-2368. DOI: 10.1103/physrevd.90.123522. URL: http://dx.doi.org/10.1103/PhysRevD.90.123522.

[40] Zachary Slepian and Daniel J. Eisenstein. "Computing the Three-Point Correlation Function of Galaxies in $\mathcal{O}(N^2)$ Time". In: *Monthly Notices of the Royal Astronomical Society* 454 (2015), p. 4142. eprint: Arxiv:1506.02040v2. URL: https://arxiv.org/abs/1506.02040.

[41] István Szapudi. "Three-Point Statistics from a New Perspective". In: 605.2 (Apr. 2004), pp. L89–L92. DOI: 10.1086/420894. arXiv: astro-ph/0404476 [astro-ph].

[42] Alessandro Benati. "Towards a joint modelling of second- and third-order clustering statistics of the Large Scale Structure". In: *Master Thesis, UNIBO* (2023).

[43] Roman Scoccimarro, H. M. P. Couchman, and Joshua A. Frieman. "The Bispectrum as a Signature of Gravitational Instability in Redshift Space". In: *The Astrophysical Journal* 517.2 (June 1999), pp. 531–540. ISSN: 1538-4357. DOI: 10.1086/307220. URL: http://dx.doi.org/10.1086/307220.

[44] Zachary Slepian and Daniel J. Eisenstein. "Modelling the large-scale redshift-space 3-point correlation function of galaxies". In: 469.2 (Aug. 2017), pp. 2059–2076. DOI: 10.1093/mnras/stx490. arXiv: 1607.03109 [astro-ph.CO].

[45] Zachary Slepian and Daniel J. Eisenstein. "On the signature of the baryon–dark matter relative velocity in the two- and three-point galaxy correlation functions". In: *Monthly Notices of the Royal Astronomical Society* 448.1 (Feb. 2015), pp. 9–26. DOI: 10.1093/mnras/stu2627. eprint: https://academic.oup.com/mnras/article-pdf/448/1/9/9376698/stu2627.pdf. URL: https://doi.org/10.1093/mnras/stu2627.

[46] J. R. Weaver et al. "COSMOS2020: A Panchromatic View of the Universe to z 10 from Two Complementary Catalogs". In: *The Astrophysical Journal Supplement Series* 258.1 (Jan. 2022), p. 11. ISSN: 1538-4365. DOI: 10.3847/1538-4365/ac3078. URL: http://dx.doi.org/10.3847/1538-4365/ac3078.

[47] Shun Saito et al. “The synthetic Emission Line COSMOS catalogue: H and [O¡scp¿ii¡/scp¿] galaxy luminosity functions and counts at 0.3 amp;lt; z amp;lt; 2.5”. In: *Monthly Notices of the Royal Astronomical Society* 494.1 (Mar. 2020), pp. 199–217. ISSN: 1365-2966. DOI: 10.1093/mnras/staa727. URL: http://dx.doi.org/10.1093/mnras/staa727.

[48] Emiliano Munari et al. “Improving fast generation of halo catalogues with higher order Lagrangian perturbation theory”. In: *Monthly Notices of the Royal Astronomical Society* 465.4 (Nov. 2016), pp. 4658–4677. ISSN: 1365-2966. DOI: 10.1093/mnras/stw3085. URL: http://dx.doi.org/10.1093/mnras/stw3085.

[49] R. J. Barlow. “Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences”. In: (1991).

[50] Antony Lewis, Anthony Challinor, and Anthony Lasenby. “Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models”. In: *The Astrophysical Journal* 538.2 (Aug. 2000), p. 473. DOI: 10.1086/309179. URL: https://dx.doi.org/10.1086/309179.